```
AAR60224
ID   AAR60224 standard; protein; 145 AA.
XX
AC   AAR60224;
XX
DT   25-MAR-2003  (revised)
DT   30-MAR-1995  (first entry)
XX
DE   Immunogenic fragment of influenza haemagglutinin.
XX
KW   Antigen; immunogen; vaccine; influenza; fusion protein; immunity;
KW   haemagglutinin; neuraminidase; flu.
XX
OS   Influenza virus.
XX
PN   WO9417826-A1.
XX
PD   18-AUG-1994.
XX
PF   01-FEB-1994;   94WO-US001149.
XX
PR   01-FEB-1993;   93US-00013415.
PR   18-AUG-1993;   93US-00108914.
PR   05-NOV-1993;   93US-00149150.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Shatzman A,  Kane J,  Scott M,  Dillon S;
XX
DR   WPI; 1994-279392/34.
XX
PT   Vaccines against multi strain influenza virus infection - protect against
PT   influenza A and B.
XX
PS   Disclosure; Page 117-118; 151pp; English.
XX
CC   A vaccine comprising an immunogenic fragment of the HA2 subunit of the
CC   influenza haemagglutinin (HA) protein from type A subtype IV and type B
CC   IV may be used for stimulating protection in animals against injection
CC   with influenza virus. The vaccine confers multi-strain immunity against
CC   strains IV A and IV B. The vaccines may be recombinantly produced,
CC   optionally as fusion proteins. This sequence is a fragment of the HA2
CC   subunit of haemagglutinin, corresponding to the amino acids at
CC   approximately 77-221 of the HA2 subunit sequence. This sequence is
CC   derived from the influenza H3 subtype. (Updated on 25-MAR-2003 to correct
CC   PN field.)
XX
SQ   Sequence 145 AA;

  Query Match             100.0%;  Score 221;  DB 2;  Length 145;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-20;
  Matches   43;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 IQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTDSEMNKLF 43
              |||||||||||||||||||||||||||||||||||||||||||
Db          1 IQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTDSEMNKLF 43
```

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : A61K 39/12, C12P 21/06, C12N 15/00, C07K 3/00, 13/00, 15/00, 17/00

A1

(11) International Publication Number: WO 94/17826

(43) International Publication Date: 18 August 1994 (18.08.94)

(21) International Application Number: PCT/US94/01149

(22) International Filing Date: 1 February 1994 (01.02.94)

(30) Priority Data:

08/013,415	1 February 1993 (01.02.93)	US
08/108,914	18 August 1993 (18.08.93)	US
08/149,150	5 November 1993 (05.11.93)	US

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM CORPORATION [US/US]; Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: SHATZMAN, Allan [US/US]; 332 Rees Drive, King of Prussia, PA 19406 (US). KANE, James [US/US]; 33 Peddrick Road, Wayne, PA 19087 (US). SCOTT, Miller [US/US]; Apartment Y-18, 103 Eagle Stream Drive, Norristown, TA 19403 (US). DILLON, Susan [US/US]; 17 Raven Drive, Chadds Ford, PA 19317 (US).

(74) Agents: BAUMEISTER, Kirk et al.; Smithkline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: VACCINAL POLYPEPTIDES

(57) Abstract

This invention provides vaccine compositions capable of conferring multi-strain immunity against influenza A and influenza B. This invention also provides methods of increasing expression and improving homogeneity of H3HA2 protein, fragments thereof, and fusion proteins containing same, as well as novel nucleotide sequences encoding these proteins and fragments.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

VACCINAL POLYPEPTIDES

Field of the Invention

The present invention relates generally to polypeptides useful in vaccine compositions and more specifically to vaccine compositions useful in providing immunity against influenza A and influenza B in an animal. The present invention also relates generally to a method of enhancing expression of polypeptides and, more specifically, to a method of enhancing influenza protein expression and homogeneity in *E. coli.*

Background of the Invention

Influenza virus infection causes acute respiratory disease in man, horses, swine and fowl, sometimes of pandemic proportions. Influenza viruses are orthomyxoviruses and, as such, have envelope virions of 80 to 120 nanometers in diameter, with two different glycoprotein spikes. Three types, A, B and C, infect humans. Type A viruses have been responsible for the majority of human epidemics in modern history, although there are also sporadic outbreaks of Type B infections. Known swine, equine, and avian viruses have mostly been Type A, although Type C viruses have also been isolated from swine.

The Type A viruses are divided into subtypes based on the antigenic properties of the hemagglutinin (HA) and neuraminidase (NA) surface glycoproteins. Within Type A, subtypes H1 ("swine flu"), H2 ("asian flu"), and H3 ("Hong Kong flu") are predominant in human infections. In swine, the predominant influenza A subtypes are H1 and H3; in horses, H3 and H7; and in avians, H5 and H7. Presently only one Type B virus has been identified, with no subtypes.

Genetic "drift" or "shift", i.e., rapid and unpredictable change in the antigen, occurs at approximately yearly intervals, and affects antigenic determinants in the HA and NA proteins. Therefore, it has not been possible to prepare a "universal" influenza virus vaccine using conventional killed or attenuated viruses, that is, a vaccine which is non-strain specific. Recently, attempts have been made to prepare such universal, or semi-universal, vaccines from reassortant viruses prepared by crossing different strains. More recently, such attempts have involved recombinant DNA techniques focusing primarily on the HA protein.

There remains a need in the art for vaccine formulations and compositions capable of inducing protective responses in animals against influenza viruses.

The expression of recombinant proteins in bacterial systems, particularly *E. coli*, is highly desirable because it can be used to produce large amounts of the desired proteins relatively inexpensively. However, high level expression of several eukaryotic proteins in *E. coli* has not been achieved for reasons including, among others, unfavorable codon usage and toxicity of the gene product [U. Brinkmann *et al.*, Gene, 85:109-114 (1989)]. Methods of overcoming these impediments to high-level expression in bacteria have been described, but are not universally applicable.

For example, Brinkmann *et al.*, cited above, described low-level expression of certain genes, such as human tissue-type plasminogen activator or gp41 of human immunodeficiency virus, which the authors attributed to the presence of the rare triplets AGA and AGG which encode arginine (Arg) in unexpectedly high amounts in the gene (3.2%). However, other eukaryotic genes, such as the NS1 gene of influenza virus, contain greater than 3% of such triplets yet express at high levels in *E. coli* [Young *et al*, Proc. Natl. Acad. Sci., 80:6105-6109 (1983)].

Another group, Spanjaard *et al.*, Nucl. Acids Res., 18(17):5031-5036 (1990) describe a translation shift in about 50% of ribosomes after tandem (double) AGA and AGG codons in cloned tRNA genes, but observed no frame shifts following single AGG or AGA codons. The authors attribute this frame shift to tRNA depletion. There also remains a need in the art for improved methods of producing vaccinal polypeptides capable of inducing protective responses in animals against influenza viruses.

Summary of the Invention

The present invention provides compositions containing, and methods for use of a protein which is capable of inducing protection in animals and avians against challenge with more than one strain of influenza Type A and influenza Type B.

Thus, one aspect of the invention provides a DNA sequence encoding a modified purified recombinant protein. The DNA sequence of the invention encodes a modified protein sequence derived from the HA2 subunit of a selected hemagglutinin (HA) protein. In one embodiment, the sequence is derived from an H3N2 subtype influenza virus. These H3N2 fusion proteins are capable of inducing T cell responses in the absence of neutralizing antibodies. In another embodiment, a DNA sequence of this invention encodes a modified protein sequence derived from the HA2 subunit from a Type B influenza virus. Still further embodiments include

DNA sequences obtained as described for the two above viruses, where the sequences are derived from other Type A influenza strains infecting animals as well as humans. Such viruses include, without limitation, Type A subtypes of H1, H2, H3, H4, H5, H6 and H7.

In another aspect, the invention provides a DNA sequence encoding a recombinant fusion protein, in which the desired Type A subtype HA2 subunit sequence or a portion thereof, is fused in frame to another protein or protein fragment capable of enhancing expression of the fusion protein. One embodiment includes the H3N2 subtype HA2 subunit sequence described above fused in frame to another protein or fragment capable of enhancing expression thereof. Another embodiment of such a fusion protein comprises a Type B HA2 sequence, described above, or a portion thereof, fused in frame to another protein or protein fragment capable of enhancing expression of the fusion protein. Additionally, other Type A subtype HA2 sequences can be similarly used. It is desirable that this fusion partner protein be an influenza protein sequence or fragment thereof.

In still another aspect, a protein encoded by a DNA sequence of the invention is provided. The protein may be a protein sequence derived from the HA2 subunit of an HA protein from a selected Type A subtype virus. Desirably the subtype virus is an H3N2. In another embodiment, the protein may be derived from the HA subunit of a Type B influenza virus. Other embodiments include H5 or H7 subtypes. Additionally, preferred embodiments include fusion proteins comprising a protein sequence derived from the HA2 subunit of an HA protein from a Type A virus, e.g., an H3N2 subtype, or from a Type B virus fused in frame to a selected influenza sequence. The proteins of this invention are particularly useful in inducing protection in mammals, especially humans, against challenge by Type B or an H3N2 subtype of influenza A. The proteins employing other Type A subtypes, e.g., H5 and H7, are useful in inducing protection in animals against influenza viruses.

In another aspect, the invention provides a method of recombinantly producing the fusion proteins of the invention, and a method of purifying the same.

In a further aspect, the invention provides a vaccine composition containing a purified protein of the invention, as described above. Such a vaccine composition may include a fusion protein of the invention. In other embodiments of the invention, the vaccine compositions contain an H3HA2 protein of the invention and other influenza antigens; a Type B HA2 protein of the invention and other influenza antigens; or both an H3HA2 protein, a BHA2 protein and other influenza antigens. In a preferred embodiment for human use, a combination vaccine of the invention will contain an H3HA2 and a BHA2 protein of the invention in

combination with influenza antigens derived from the other Type A influenza virus subtypes, H1 and H2. An embodiment for use in animals may contain an H5HA2 or H7HA2 protein, among others.

A further aspect of this invention is a method for inducing in an animal protection against influenza Type A, influenza Type B, influenza Type C, or combinations thereof, which comprises internally administering to the animal an effective immunogenic amount of a vaccine composition of the present invention.

Still a further aspect of this invention is a method for inducing in an animal protection against multiple strains of influenza Types A and B which comprises internally administering to the animal an effective immunogenic amount of a vaccine composition of the present invention.

In another aspect, the present invention provides a method of enhancing in *E. coli* the expression of influenza vaccinal proteins characterized by a naturally-occurring amino acid pattern comprising Arg-Arg-Xaa-Xaa-Arg [SEQ ID NO:8]. In this pattern, Arg is arginine, Xaa is any amino acid, and at least one of the arginines in the naturally-occurring sequence is encoded by the rare nucleic acid triplets AGG or AGA.

In one embodiment, the method of the invention involves mutating one or more of these AGG or AGA codons to a preferred arginine codon and expressing the mutated sequence in *E. coli*. Surprisingly, it has been found that this modification, which does not result in a change in the encoded amino acid sequence, can increase the expression and homogeneity of an influenza protein in *E. coli* significantly.

In another embodiment, the method of this invention involves increasing the expression of the above-identified proteins by inserting into the host cell tRNA molecules capable of translating the native rare arginine codons. Thus, the *E. coli* host cells are modified such that they are capable of efficiently translating the rare, native arginine codons.

In another aspect, the present invention provides novel nucleic acid sequences of influenza proteins which contain the nucleotide sequence CGn-CGn-Xaa-Xaa-CGn, where n represents a nucleotide selected from the group consisting of T, C, A or G [SEQ ID NO:9], in place of the native nucleotide sequence AGr-AGr-Xaa-Xaa-AGr, where r represents the nucleotides A or G [SEQ ID NO:10]. When expressed in *E. coli*, these sequences result in increased expression of the encoded protein as compared to the native sequence.

In still another aspect, the invention provides the novel modified nucleic acid sequences described above fused in the same reading frame to another

DNA sequence encoding a polypeptide or protein, i.e., a fusion partner, which may further enhance the expression of, or immunogenicity of, the encoded influenza protein. It is desirable that the fusion partner be an influenza protein sequence or fragment thereof.

Other aspects and advantages of the present invention are described further in the following detailed description of the preferred embodiments thereof.

Brief Description of the Drawings

Fig. 1 illustrates the nucleic acid sequences of the HA2 portions of (a) A/Udorn [SEQ ID NO: 1], (b) A/Victoria [SEQ ID NO: 3], (c) A/PR/8/34 [SEQ ID NO: 5], and (d) a consensus sequence [SEQ ID NO: 7]. Dashes indicate the same nucleotide as the consensus sequence. Different nucleotides from that of the consensus sequence are reported in lower case letters. Dots indicate no corresponding nucleotide when compared to the consensus sequence.

Fig. 2 illustrates the nucleic acid and amino acid sequences of H3C13, $NS1_{(1-81)}H3HA2_{(1-221)}$ fusion protein [SEQ ID NO: 9 & 10], with the mutant nucleic acid sequences of H3C13mut5855 [SEQ ID NO: 58] illustrated above the sequence of the unmodified H3HA2 portion.

Fig. 3 illustrates the nucleic acid and amino acid sequences of the $NS1_{(1-81)}H3HA2_{(77-221)}$ fusion protein [SEQ ID NO: 11 & 12].

Fig. 4 illustrates the nucleic acid and amino acid sequences of the Type B fusion protein, $NS1_{(1-42)}HA2_{(41-223)}$. [SEQ ID NO: 13 & 14].

Fig. 5 illustrates the pOTS208NS1BLmut2 vector nucleic acid sequences [SEQ ID NO: 54] encoding the amino acid sequences [SEQ ID NO: 55] of the mutant $NS_{(1-81)}BLHA2_{(1-223)}$(met-leu) fusion protein, with the nucleic acid sequences of the coding region $NS_{(1-81)}BLHA2_{(1-223)}$ [SEQ ID NO: 56] and native amino acid sequences [SEQ ID NO: 57], which include a Met in amino acid position 98, illustrated above the modified BLHA2 sequences.

Fig. 6 illustrates the nucleic acid [SEQ ID NO:17] and amino acid [SEQ ID NO:18] sequences of the H1N1 fusion protein, $NS1_{(1-81)}HA2_{(65-222)}$, also known as flu D.

Fig. 7 illustrates the naturally-occurring nucleic acid sequence [SEQ ID NO:1] and corresponding amino acid sequence [SEQ ID NO:2] of the HA2 portion of the H3N2 virus, A/Udorn.

Fig. 8 illustrates the naturally-occurring nucleic acid sequence [SEQ ID NO:3] and corresponding amino acid sequence [SEQ ID NO:60] of the HA2 portion of the H3N2 virus, A/Victoria.

Detailed Description of the Invention

The present invention provides novel proteins, DNA sequences, pharmaceutical vaccine compositions, and methods of use thereof for conferring protection in vaccinated mammals against one strain, or desirably multiple strains, of influenza viruses. The proteins and vaccine compositions of the present invention demonstrate the ability to stimulate or produce a protective immune response which is capable of recognizing an influenza virus or influenza virus-infected cells and protecting the vaccinated mammal against disease caused thereby. This protective response is desirably a T cell response, produced in the substantial absence of vaccine-induced neutralizing antibody.

While the proteins and DNA sequences specifically described herein are directed to the H3HA2 and BHA2 sequences originating from viral strains to which humans are susceptible, it is expected that similar sequences and molecules can be prepared for veterinary applications. For example, selected HA2 sequences obtained from Type A viral strains, e.g., H5HA2, H7HA2 and other strains of interest may be obtained following the teachings described herein for the exemplified H3HA2 and BHA2 sequences. One of skill in the art should understand that this invention is not limited to the exemplified protein and DNA sequences, even though the following disclosure is limited to the two latter sequences for simplicity. Such additional viral HA2 subunits are expected to share the biological characteristics of the exemplified sequences.

Thus, this invention provides a protein or fragment thereof characterized by an amino acid sequence derived from the HA2 subunit of an HA protein, e.g., from a H3N2 subtype virus. As used herein, a "fragment" of the HA2 subunit is an amino acid sequence derived from the HA2 subunit which is characterized by having an immunogenic determinant of the HA2 subunit. Such a fragment is desirably at least about 8 amino acids in length.

The H3 proteins of the invention are capable of inducing T helper cells, particularly cytotoxic T lymphocytes, in the absence of neutralizing antibodies. Among H3N2 subtype strains of influenza A include A/Udorn and A/Victoria viruses. Other H3N2 virus strains of influenza A may also produce HA proteins for use in vaccine compositions according to this invention. Fig. 1 compares the nucleic acid sequences of the HA2 portions of the A/Udorn [SEQ ID NO: 1] and A/Victoria [SEQ ID NO: 3] strains with the nucleic acid sequence of an H1N1 subtype virus, A/PR/8/34 [SEQ ID NO: 5]. A consensus sequence [SEQ ID NO: 7] was computer generated, and may likewise be useful in producing proteins

according to this invention. This consensus sequence [SEQ ID NO: 7] can be constructed by a commercially available computerized sequence analysis program, such as Genetics Computers Group [University of Wisconsin].

Proteins according to this invention may include unfused HA2 subunits of the influenza A viruses, particularly H3N2 subtype. For example, in one embodiment, a protein of the invention contains amino acids 1-221 of a selected H3HA2 subunit. In another embodiment, a protein of the invention contains amino acids 77-221 of the H3HA2 subunit. Other fragments of this HA2 amino acid sequence characterized by the ability to stimulate similar immunological activity in an immunized animal are also encompassed by this invention.

Proteins of this invention also include fusion proteins comprising a protein sequence derived from the HA2 subunit of an HA protein from a Type A virus, e.g., an H3N2 subtype virus, fused in frame to another protein or protein fragment capable of enhancing expression of the fusion protein. It is desirable that this fusion "partner" protein be an influenza protein sequence or fragment thereof derived from the same or another strain of influenza virus as the HA protein or protein fragment. Preferably, this fusion partner protein is all or a portion of the influenza virus NS1 protein or an HA2 subunit protein.

In the embodiments exemplified herein, the NS1 portion of the fusion protein is derived from an H1N1 subtype virus, A/PR/8/34. For example, in one embodiment, the NS1 portion may comprise amino acid residues 1 to 42 of H1NS1. In another embodiment the NS1 portion may comprise amino acid residues 1 to 81 of the selected virus. The HA2 fragment may alternatively be fused to a portion of the NS1 peptide derived from a selected Type A virus, e.g., an H3 subtype virus (H3HA2), or a Type B (BHA2) virus.

However, other non-influenza fusion proteins may also produce desirable fusion proteins with the H3N2, or other Type A, or Type B protein or portion thereof. Thus, in still another alternative embodiment, as discussed below, the HA2 fragment may be fused to any peptide capable of enhancing its expression in the host cell selected. One of skill in the art may readily select a fusion "partner" protein or fragment taking into account the desired host cell and utilizing the teachings herein. The fusion proteins of the present invention are not limited by the selection of the "partner" protein or fragment to which the HA2 fragment is fused.

In yet another embodiment, the present invention provides a modified protein containing a portion of the HA2 subunit of a Type B influenza virus. Currently, the preferred human virus strain is B/Lee/40. However, the vaccinal proteins of this invention are not limited to this Type B strain, and other

strains infecting other species, or other as yet unidentified Type B virus strains, may be used to produce the HA2 protein. These Type B HA2 proteins may be fused to a fusion "partner" protein or protein fragment, as described above for the H3HA2 proteins of this invention, or remain unfused.

In the construction of a fusion protein according to this invention, a linker sequence may optionally be inserted between the two fused sequences, i.e., between the NS1 portion and the HA2 portion. This optional linker may provide space between the two linked sequences. Alternatively, this linker sequence may encode, if desired, a polypeptide which is selectively cleavable or digestible by conventional chemical or enzymatic methods. For example, the selected cleavage site may be an enzymatic cleavage site, including sites for cleavage by a proteolytic enzyme, such as enterokinase, factor Xa, trypsin, collagenase, and thrombin. Alternatively, the cleavage site in the linker may be a site capable of being cleaved upon exposure to a selected chemical, e.g., cyanogen bromide or hydroxylamine. The cleavage site, if inserted into a linker useful in the fusion sequences of this invention, does not limit this invention. Any desired cleavage site, of which many are known in the art, may be used for this purpose.

A presently preferred example of an H3 fusion protein of this invention is $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10], which comprises the first 81 amino acids of NS1 fused to amino acids 1 to 221 of the H3HA2 subunit (amino acids 1-221). (Fig. 2) Another exemplary fusion protein, $NS1_{(1-81)}H3HA2_{(77-221)}$ [SEQ ID NO: 12], comprises the first 81 amino acids of NS1 fused to amino acids 77 to 221 of the truncated H3HA2 subunit. (Fig. 3)

A present preferred example of a Type B fusion protein of this invention is $NS1_{(1-42)}BHA2_{(41-223)}$ [SEQ ID NO: 14], which comprises the first 42 amino acids of NS1 fused to amino acids 41 to 223 of the truncated BHA2 subunit. (Fig. 4) Another fusion protein of this invention is $NS1_{(1-81)}BHA2_{(1-223)}$ [SEQ ID NO: 57], which contains the first 81 amino acids of NS1 fused to amino acids 1 to 223 of the BHA2 subunit. (Fig. 5) Another preferred fusion protein of the invention is $NS1_{(1-81)}BHA2_{(1-223)}$(met-leu) SEQ ID NO: 55, which contains the same amino acid sequence as $NS1_{(1-81)}BHA2_{(1-223)}$, with the exception that the internal methionine residue at position 98 of the fusion protein has been changed to a leucine. (Fig. 5)

These proteins, fusion proteins, and similar proteins encoded by the below-described DNA sequences are referred to collectively herein as H3HA2 proteins.

The $NS1_{(1-81)}H3HA2_{(1-221)}$ protein [SEQ ID NO: 10] of the invention has a three-dimensional structure which is substantially similar to that of the $NS1_{(1-81)}HA2_{(1-222)}$ protein [SEQ ID NO: 16] derived from the H1N1 subtype virus (C13). However, the amino acid sequence of the $NS1_{(1-81)}H3HA2_{(1-221)}$ protein [SEQ ID NO: 10] has only approximately 50% homology with the amino acid sequence of C13 protein [SEQ ID NO: 16]. Additionally, as illustrated in Fig. 1, the nucleic acid sequence of the $H3HA2_{1-221}$ protein derived from A/Udorn (nucleotides 25-560 from that virus) [SEQ ID NO: 1] has only approximately 60% homology with the nucleic acid sequence of the $H1HA2_{1-222}$ protein derived from strain A/PR/8/34 (nucleotides 1872-2407 from A/PR/8/34) [SEQ ID NO: 5]. However, the nucleic acid sequence of $H3HA2_{1-221}$ from A/Udorn (nucleotides 1-499 of A/Udorn) [SEQ ID NO: 1] has approximately 99% homology with the nucleic acid sequence of $H3HA2_{1-221}$ from A/Victoria/H3/75 (nucleotides 1226-1725 of A/Victoria) [SEQ ID NO: 3] [Fiers et al, Cell, 19:683-696 (1980)].

Analogs of the HA2 peptides from a Type A virus, e.g., an H3, or Type B viruses, included within the definition of this invention, include truncated polypeptides (including fragments) and HA2 polypeptides, e.g. mutants that retain the epitopes and thus the biological activity of HA2. It is anticipated that, because the NS1 portion of the fusion peptide provides a means of expressing the protein at high levels and does not appear to play as significant a role in the immunological responses to the HA2 fusion proteins as does the HA2 portion, any number of analogs of this fusion partner can be made.

Typically, the analogs of the HA2 peptides and/or the fusion partner differ by only 1 to about 4 codon changes. Other examples of analogs include polypeptides with minor amino acid variations from the natural amino acid sequence of HA2; in particular, conservative amino acid replacements. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are generally divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) non-polar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid will not have a

significant effect on its activity, especially if the replacement does not involve an amino acid at an epitope of the HA2 polypeptide. The construction of such analogs, given the description herein and conventional methods of protein modification known to one of skill in the art, are believed to be encompassed by this invention.

Currently, it is theorized that the HA2 portion of the fusion peptide (e.g., $H3HA2_{1-221}$, $H3HA2_{77-221}$ and $BHA2_{41-223}$) confers the majority of the necessary epitopes for antibody binding or T cell (particularly CTL) targeting. Once these epitope sequences are precisely identified, portions of the HA2 sequence which are not part of these epitopes may be altered without significantly affecting the bioactivity of the fusion protein.

The present invention also encompasses DNA sequences of this invention encoding the above-described proteins and fusion proteins, the sequences characterized by having an immunogenic determinant of a modified HA2 subunit of an HA protein, derived from a Type A virus, e.g., an H3 subtype, or Type B virus. Other DNA sequences of this invention encode such HA2 subunits, optionally fused to a DNA sequence encoding a protein or peptide which is capable of enhancing expression of the protein in a selected host cell. For example, the consensus sequence illustrated in Fig. 1(d) may provide a source of HA2 DNA. The currently preferred embodiment provides a DNA sequence encoding a Type A virus, e.g., an H3 or Type B HA2 protein or fragment thereof fused in frame to a DNA sequence encoding a portion of the nonstructural influenza protein 1 (NS1).

Coding sequences for the HA2, NS1, and other viral proteins of influenza virus can be prepared synthetically or can be derived from viral RNA or from available cDNA-containing plasmids by known techniques. For example, in addition to the above-cited references, a DNA coding sequence for HA from the A/Japan/305/57 strain was cloned, sequenced and reported by Gething et al, Nature, 287:301-306 (1980). An HA coding sequence for strain A/NT/60/68 was cloned as reported by Sleigh et al, and by Both et al, in Developments in Cell Biology, Elsevier Science Publishing Co., pages 69-79 and 81-89, respectively, (1980). An HA coding sequence for strain A/WSN/33 was cloned as reported by Davis et al, Gene, 10:205-218 (1980); and by Hiti et al, Virology, 111:113-124 (1981). An HA coding sequence for fowl plague virus was cloned as reported by Porter et al and by Emtage et al, both in Developments in Cell Biology, cited above, at pages 39-49 and 157-168. Also, influenza viruses, including other strains, subtypes, and types are available from clinical specimens and from public depositories, such as the American Type Culture Collection (ATCC), Rockville, Maryland, U.S.A.

Allelic variations (naturally-occurring base changes in the species population which may or may not result in an amino acid change) of DNA sequences encoding the H3HA2 or BHA2 protein sequences are also included in the present invention, as well as analogs or derivatives thereof. Similarly, DNA sequences which code for H3 or other Type A or Type B HA2 proteins of the invention but which differ in codon sequence due to the degeneracies of the genetic code or variations in the DNA sequence encoding H3HA2, other Type A or BHA2 proteins which are caused by point mutations or by induced modifications to enhance the activity, half-life or production of the peptide encoded thereby are also encompassed in the invention. Suitably, this invention provides certain silent mutations to the coding sequences for $NS1_{(1-81)}H3HA2_{(1-221)}$, which have been found to increase expression yields. See Fig. 2. Further, the $NS1_{(1-81)}BHA2_{(1-223)}$(met-leu)-encoding sequence, BC13mut2, in addition to modifying the codon encoding amino acid position 98 of the fusion protein (position 17 of the HA2 portion), contains a number of silent modifications designed to increase protein expression. See Fig. 5.

Also covered by this invention are DNA sequences which hybridize under stringent conditions with the DNA sequences encoding the HA2 subunit proteins, e.g., H3HA2 or BHA2 proteins, of this invention. DNA sequences which hybridize under non-stringent conditions with the disclosed sequences, but which encode proteins or fragments retaining the biological activities of the H3HA2 or BHA2 proteins, are also included in this invention. Typical conditions for stringent or non-stringent hybridization are known to those of skill in the art. [See, e.g., Sambrook et al, Molecular Cloning. A Laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory, NY (1989)].

The fusion proteins of the invention may be prepared by conventional genetic engineering and recombinant techniques known to those of skill in the art. Similarly, the proteins may be purified from expression in host cell or vector systems by conventional means.

Preferably, however, the recombinantly-produced fusion proteins of the invention are purified as described herein. Generally, method of purification involves (step 1) the isolation of the proteins, (step 2) enzymatic digestion and extraction, (step 3) urea extraction, (step 4) solubilization, reduction, and DEAE chromatography, (step 5) reverse phase chromatography, (step 6) precipitation, and (step 7) desalting and preparation of the final product. More specifically, the host cells containing the fusion proteins are disrupted, either chemically or by mechanical means. Preferably the cells are lysed by osmotic shock. Following

centrifugation, the resulting pellet (P1) is subjected to nuclease digestion extraction and centrifuged to yield pellet 2 (P2). A second extraction step is then performed using urea (pH 6) and the mixture centrifuged to yield pellet 3 (P3). P3 is then solubilized and reduced. Preferably, solubilization is performed using urea at pH 12.5 and reduction is via DTT DEAE chromatography followed by SDS elution. The resulting DEAE product is further reduced, preferably using DTT, and subjected to reverse phase chromatography. The reverse phase product is then precipitated by adjusting to pH 6 and centrifuged. The precipitated product is resolubilized, preferably with urea at pH 12.5, and subjected to G25 chromatography. The resulting G25 product is then filtered (e.g. with a 0.2 micron filter) to yield the final product. Further details of this method are provided in Example 17 below.

Systems for cloning and expression of the vaccinal polypeptide of this invention in various microorganisms and cells, including, for example, E. coli, Bacillus, Streptomyces, Saccharomyces, mammalian and insect cells, are known and available from private and public laboratories and depositories and from commercial vendors. The preferred host is E. coli because it can be used to produce large amounts of desired proteins safely and cheaply. To circumvent the requirement of ampicillin for plasmid selection in production fermentations, a desirable method of production employs an alternative expression system in which the β-lactamase coding sequence is wholly or partially replaced by a coding sequence for an alternative selectable marker such as, for example, kanamycin or chloramphenicol.

Thus, the polypeptide employed in the presently preferred embodiment is preferably expressed in E. coli. A suitable strain, LW14, has the following genotype: galE::Tn10λCI857 bio- uvrB-; phenotypically, strain LW14 requires biotin for growth, is sensitive to UV light and DNA damaging agents, and cannot use galactose as a carbon source. Construction of this strain is described in the examples below.

To aid in expression of the H3 or other Type A subunit or Type B HA2 peptides or fusion protein described above, these protein sequences or fragments thereof may also be fused to a polypeptide capable of enhancing expression of these fragments in the selected host system. Ordinarily, such a peptide would contain a leader sequence fragment that provides for secretion of the Type A subunit fragment, e.g., the H3HA2 fragment, or Type B HA2 fragment in the host cell. The leader sequence fragment typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. There may be processing sites encoded between the leader sequence

and the Type A subtype or Type B HA2 fragment that can be cleaved either in vivo or in vitro. Alternatively, a promoter sequence may be linked directly with the DNA molecule encoding the HA2 fragment. Such polypeptides, promoter and leader sequences are known to those of skill in the art and may be readily selected for expression in the selected host.

Construction of expression systems, including expression vectors and transformed host cells are thus within the art. See, generally, methods described in standard texts, such as Sambrook et al, Molecular Cloning A Laboratory Manual, 2d edit., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989). The present invention is therefore not limited to any particular expression system or vector, nor to any particular purification process from cell lysates or cell medium.

The proteins and fusion proteins of this invention may be employed in vaccine compositions. Pharmaceutical vaccine compositions of this invention, therefore, contain an effective immunogenic amount of a selected HA2 protein, e.g., H3HA2 or BHA2 protein, of the invention in admixture with a suitable adjuvant in a nontoxic and sterile pharmaceutically acceptable carrier.

Suitable carriers for vaccine use are well known to those of skill in the art. However, exemplary carriers include sterile saline, lactose, sucrose, calcium phosphate, gelatin, dextrin, agar, pectin, peanut oil, olive oil, sesame oil, squalene, and water. Additionally, the carrier or diluent may include a time delay material, such as glyceryl monostearate or glyceryl distearate alone or with a wax. Optionally, suitable chemical stabilizers may be used to improve the stability of the pharmaceutical preparation. Suitable chemical stabilizers are well known to those of skill in the art and include, for example, citric acid and other agents to adjust pH, chelating or sequestering agents, and antioxidants.

While any aluminum adjuvant may be used in the vaccine compositions of this invention, two desirable adjuvants are available commercially, i.e., REHSORPTAR™ adjuvant [Armour Pharmaceuticals, Kankakee, IL] and REHYDRAGEL™ adjuvant [Reheis Chemical Co., Berkeley Heights, NJ]. These products are aluminum hydroxide gels which contain approximately 2% w/v Al_2O_3, which is equivalent to approximately 10.6 mg/ml Al^{+3}.

Vaccine compositions of this invention may employ an immunogenic amount of a purified recombinant protein as described above. A preferred embodiment of the vaccine of the invention is composed of an aqueous suspension or solution containing the recombinant HA2 protein molecule, e.g., H3HA2 or BHA2, together with an adjuvant, preferably an aluminum, most preferably aluminum hydroxide, buffered at physiological pH, in a form ready for injection. A

preferred protein for use in these vaccine compositions includes a protein comprising amino acid residues 1 to 81 from NS1 fused to C-terminal amino acid residues 1-221 from the hemagglutinin subunit 2 (HA2) from influenza A, subtype H3N2. Another preferred vaccine composition of this invention employs a purified recombinant protein made up of amino acid residues 1 to 81 from NS1 fused to amino acid residues 77-221 of the HA2 from influenza A, subtype H3N2. Still another preferred vaccine composition of this invention employs a purified recombinant protein made up of amino acid residues 1 to 42 fused to amino acid residues 41-223 of the HA2 from influenza B.

Vaccine compositions of the invention may also employ an immunogenic amount of a recombinant protein of the invention in combination with other influenza antigens. Suitable influenza antigens for combination in a vaccine composition with the proteins of this invention may be derived from Type A, H1 subtype viruses and may include the recombinant fusion proteins described in detail in copending U. S. Patent Application Ser. No. 07/387,200, filed July 28, 1989 and its corresponding European Patent Application No. 366, 238, published May 2, 1990; and in co-pending U. S. Patent Application Ser. No. 07/387,558, filed July 28, 1989 and its corresponding European Patent Application No. 366,239, published May 2, 1990. The C13 protein ($NS1_{(1-81)}HA2_{(1-222)}$) [SEQ ID NO: 15 & 16], D protein ($NS1_{(1-81)}HA2_{(65-222)}$) [SEQ ID NO: 17 & 18] and other fusion proteins derived from the H1N1 influenza virus subtype and the recombinant expression and purification thereof are disclosed in detail in these applications, and in the parent applications identified in this application, all of which are incorporated by reference herein.

More specifically, suitable H1 subtype immunogenic proteins include C13 ($NS1_{(1-81)}$-D-L-S-R-$HA2_{(1-222)}$) [SEQ ID NO: 15 & 16], D ($NS1_{(1-81)}$-Q-I-P-$HA2_{(65-222)}$) [SEQ ID NO: 17 & 18], C13 short ($NS1_{(1-42)}$-M-D-L-S-R-$HA2_{(1-222)}$) [SEQ ID NO: 19 & 20], D short ($NS1_{(1-42)}$-M-D-H-M-L-T-S-T-R-S-$HA2_{(66-222)}$) [SEQ ID NO: 21 & 22], A ($NS1_{(1-81)}$-Q-I-P-$HA2_{(69-222)}$) [SEQ ID NO: 23 & 24], C ($NS1_{(1-81)}$-Q-I-P-$HA2_{(81-222)}$) [SEQ ID NO: 25 & 26], ΔD ($NS1_{(1-81)}HA2_{(150-222)}$) [SEQ ID NO: 27], Δ13 ($NS1_{(1-81)}$-D-L-S-R-$HA2_{(1-70)}$-S-C-L-T-A-Y-H-R) [SEQ ID NO: 28], M ($NS1_{(1-81)}$-Q-I-P-$HA2_{(65-196)}$-G-G-S-Y-S-M-E-H-F-R-W-G-K-P-V) [SEQ ID NO: 29], ΔM ($NS1_{(1-81)}$-Q-I-P-$HA2_{(65-196)}$-G-G-S-Y-S-M-L-V-N) [SEQ ID NO: 30], ΔM+ ($NS1_{(1-81)}$-Q-I-P-$HA2_{(65-200)}$-L-V-L-L) [SEQ ID NO: 31 & 32]. These H1N1 fusion proteins are described in published European Patent Application 366,238 and in copending U.S. Patent Application Ser. No. 07/751,896. Other suitable H1 proteins consist of

unfused polypeptides, such as H1HA2$_{66-222}$ [SEQ ID NO: 33 & 34] which is disclosed in co-pending U. S. Patent Application Ser. No. 07/751,898, incorporated herein by reference. Thus, one desirable combination vaccine to provide protection against Type A influenza contains NS1$_{(1-81)}$H3HA2$_{(1-221)}$ protein [SEQ ID NO: 9 & 10] of the invention, one or more proteins derived from subtype H1N1 as described above, and an aluminum adjuvant.

Preferably, a combination vaccine of the invention will contain an immunogenic amount of the H3 fusion protein of the invention in combination with immunogenic amounts of influenza antigens derived from the other Type A influenza virus subtypes, including among others, H1, H2, H3, H4, H5, H6; and H7, as well as a Type B fusion protein of the invention.

A currently preferred combination vaccine of the invention contains the H3 subtype fusion protein NS1$_{(1-81)}$H3HA2$_{(1-221)}$ [SEQ ID NO: 10], the B subtype fusion protein NS1$_{(1-81)}$BHA2$_{(1-223)}$(met-leu) [SEQ ID NO: 55], and the H1 subtype fusion protein NS1$_{(1-81)}$HA2$_{(65-222)}$ [SEQ ID NO: 18]. Studies have shown that such a combination vaccine is protective against challenge with H1, H3 and Type B influenza viruses in mice.

Other preferred combination vaccines would include the NS1$_{(1-81)}$H3HA2$_{(77-221)}$ protein [SEQ ID NO: 12] or the NS1$_{(1-81)}$BHA2$_{(1-223)}$ [SEQ ID NO: 57] in combination with one or more additional influenza antigens derived from the type or subtype influenza viruses described above. These combination vaccines will protect against influenza infections caused by both Type A and Type B influenza viruses. Still other combination vaccine compositions will employ other proteins described herein.

The compositions of the present invention are advantageously made up in a dose unit form adapted for the desired mode of administration. Each unit will contain, at a minimum, a predetermined quantity of the selected HA2 subunit protein, e.g., H3HA2 protein and/or BHA2 protein, and adjuvant calculated to produce the desired therapeutic effect in optional association with a pharmaceutical diluent, carrier or vehicle.

Dosage protocol can be optimized in accordance with standard vaccination practices. Typically, the vaccine will be administered intramuscularly, although other routes of administration may be used, such as intradermal. It is expected that an effective immunogenic amount of a protein, fusion protein or combination of proteins of this invention for average adult humans is in the range of 1 to 1000 micrograms. Another desirable immunogenic amount ranges between 50

to 500 micrograms. Most preferably, the proteins of the invention are in admixture with the same amount or more adjuvant to form a vaccine composition.

While the proteins described herein have been particularly developed for use in humans (e.g., the H3HA2 and BHA2 sequences), it is expected that due to species cross-reactivity, these vaccines will be useful in other animals, particularly swine. Additionally, similar molecules can be prepared for equine and avian veterinary applications utilizing the HA2 proteins from other strains to which animals are susceptible. Combination vaccines for use in swine would preferably include protections against both H1 and H3 viruses. Combination vaccines for use in equine would preferably include protection against H3 and H7 viruses. Combination vaccines for use in avian species would preferably confer protection against H5 and H7 viruses. Appropriate dosages can be determined by one skilled in veterinary medicine.

It will be understood, however, that the specific effective immunogenic amount for any particular patient will depend upon a variety of factors including the age, general health, sex, and diet of the vaccinee; the species of the vaccinee; the time of administration; the route of administration; interactions with any other drugs being administered; and the degree of protection being sought.

The vaccine can be administered initially in late summer or early fall and can be readministered two to six weeks later, if desirable, or periodically as immunity wanes, for example, every two to five years. Of course, as stated above, the administration can be repeated at suitable intervals if necessary or desirable.

The present invention provides methods for producing enhanced expression and improved homogeneity of influenza viral proteins and polypeptides in *E. coli*. Also provided are novel modified nucleotide sequences which encode these influenza proteins and are useful in the methods of production.

Preferably, the influenza proteins or polypeptides produced according to the invention include the complete HA2 protein of the hemagglutinin antigen (HA) of a selected H3N2 influenza virus, a complete HA protein of an H3HA2 virus, fragments thereof, and fusion proteins containing the complete H3HA2 protein or desired fragments thereof fused in the same reading frame with a selected fusion partner polypeptide or protein. These proteins are characterized by having the native amino acid sequence pattern described above.

By the term "fragment" is meant a subunit of HA, or a span of contiguous amino acids from the complete protein capable of stimulating an antigenic or protective immunogenic response in an animal. A fragment may contain at least about 8 amino acids from the selected influenza protein, and can

contain up to the number of amino acids which make up the entire protein. When the term 'fragment' is used herein to modify a nucleotide sequence, it refers to nucleotide sequences which encode the above-defined amino acid fragments.

Native (or naturally-occurring) nucleotide sequences which encode certain influenza proteins are characterized by a nucleotide sequence pattern encoding the fragment Arg-Arg-Xaa-Xaa-Arg [SEQ ID NO:61]. Arg represents arginine and Xaa represents any amino acid in this formula. Hereafter, this five amino acid sequence is referred to as Formula I.

Formula I sequences are typically encoded by native nucleotide sequences of the formula of codons AGr-AGr-Xaa-Xaa-AGr, where r represents the nucleotides A or G and Xaa represent any codon [SEQ ID NO:63]. Hereafter, this five codon nucleotide sequence is referred to as Formula II. Specifically, the native nucleic acid sequence encoding a subtype H3N2 influenza virus protein, fusion protein, or a fragment or subunit thereof, specifically the HA2 portions of H3N2 virus strains, is characterized by a Formula II sequence.

Among H3N2 subtype strains of influenza A characterized by this nucleotide fragment Formula II include the A/Udorn and A/Victoria viruses. Figs. 7 and 8 provide the native nucleic acid sequences of the HA2 portions of the A/Udorn [SEQ ID NO: 1] and A/Victoria [SEQ ID NO: 3] strains. Other H3N2 virus strains of influenza A may also provide native nucleotide sequences containing Formula II, which sequences are susceptible to the modifications described herein.

Additional examples of native nucleotide sequences encoding proteins whose expression may be enhanced according to this invention are those native sequences which encode certain fragments of influenza proteins including the fragment spanning amino acids 1 to about amino acids 221 of H3HA2 [Fig. 7 SEQ ID NO:2 and Fig. 8 SEQ ID NO:3]; the fragment spanning from about amino acid 77 to about amino acid 221 [Fig. 7 SEQ ID NO:69 and Fig. 8 SEQ ID NO:70], or other desirable fragments. Other desirable fragments of this H3HA2 amino acid sequence include those characterized by the ability to stimulate immunological activity in an immunized animal similar to that stimulated by use of the entire 221 amino acid sequence of H3HA2.

Nucleotide sequences encoding fusion proteins which contain fragments of the native nucleotide sequences encoding these influenza proteins or subunits, e.g., the fusion protein $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO:10], can also be characterized by the Formula II nucleotide sequence. Thus these fusion proteins are also desirable for enhanced expression according to the method of this invention.

The inventors have discovered that when native nucleotide sequences of influenza proteins, which sequences comprise Formula II, are expressed in *E. coli*, a frame shift of one nucleotide after the third triplet in Formula II in the native sequence occurs, resulting in the increased translation of truncated proteins. It has been surprisingly found that by application of a method of the present invention, the expression and homogeneity of the influenza protein is increased significantly.

The methods of this invention involve enhancing the expression of proteins characterized by the amino acid pattern of Formula I, which proteins have a native nucleotide sequence of Formula II. According to one embodiment of the method of this invention, a native nucleotide sequence encoding a selected influenza protein or fragment, which sequence comprises Formula II, is modified by mutating one or more of the rare AGG or AGA arginine codons of Formula II to a preferred Arg codon. A preferred arginine codon for use in replacing a native AGA or AGG codon according to this invention is defined herein by the codons CGT, CGG, CGA and CGC. Of these codons, CGT and CGC are currently the most preferred. The modified influenza protein-encoding nucleotide sequence is then expressed in an *E. coli* expression system, resulting in enhanced expression in comparison to that obtained by expression of the native nucleotide sequence encoding the same protein in the same expression system.

The enhanced protein expression occurs even though the mutation does not result in a change in the encoded amino acid sequence of the protein. By the terms 'enhanced expression' or 'enhanced protein expression' is meant an expression level of at least 40% higher than the expression level of the protein encoded by the native, non-mutated nucleotide sequence comprising Formula II, when expressed in *E. coli*.

While not wishing to be bound by theory, the inventors believe that the enhanced expression levels are obtained because the silent mutation of the AGA or AGG to a preferred arginine codon in Formula II eliminates the frame shift mutation found in the unmutated nucleotides encoding these proteins, thus substantially reducing the production of truncated messages (proteins). It is believed that the resulting influenza proteins are more homogeneous when expressed in an *E. coli* expression system according to this invention.

In a second embodiment of the method of the invention, the expression of the proteins containing arginines encoded by the rare codons AGG and AGA (i.e. proteins encoded by amino acid and nucleotide sequences characterized by Formulae I and II) can be increased by inserting into the host in which expression is desired one or more genes for tRNA molecules which are

capable of properly translating the AGG and AGG arginine codons. Preferably the host cells are *E. coli.*

This method can be accomplished as follows. A gene for a tRNA molecule described above can be selected from among known gene sequences. The genes and tRNA molecules which can translate the rare Arg codons identified above are known and readily available to one of skill in the art. See, e.g., [P. Saxena and J. Walker, J. Bacteriol., 174(6):1956-1964 (Mar. 1992)].

According to conventional techniques, these genes may be placed on a plasmid which will increase the copy number of these genes and therefore the tRNA molecules encoded by these genes. Alternatively, these sequences can be genetically engineered and placed on the host cell chromosome behind an appropriate promoter element in such a manner that the effective concentration of these tRNA molecules is increased inside the cell. Conventional texts describe the techniques useful in this method [See, e.g., Sambrook et al., Molecular Cloning. A Laboratory Manual. 2d edition, Cold Spring Harbor, New York (1989)].

The insertion of the tRNA genes into the host cell expressing the protein increases the concentration of these tRNA molecules inside the host cells which are naturally deficient for these tRNA molecules. This allows the host cells to translate these rare arginine codons in an efficient manner, eliminating the production of the truncated or lower molecular weight species of the fusion protein observed in the unmodified host cell. Thus, this method may be used to increase expression of a protein in host cells lacking sufficient amounts of the appropriate tRNA to permit efficient expression of the protein. Use of this method obviates the need to modify the sequences encoding the selected protein, and thus provides an alternative method to the first embodiment described above.

As another aspect of this invention novel modified nucleotide sequences are provided, which in *E. coli* expression systems, can be employed to produce the encoded influenza proteins, subunits, fragments and fusion proteins described above according to the first embodiment of the method of this invention. The proteins encoded by these nucleotides are produced at levels of expression enhanced over that of the native sequences, by about forty percent or more. The novel nucleotide sequences of the invention are characterized by comprising the nucleotide sequence CGn-CGn-Xaa-Xaa-CGn, where n represents a nucleotide selected from the group consisting of T, C, A or G [SEQ ID NO:62], in place of the Formula II fragment in the native nucleotide sequence encoding the selected influenza protein or fragment. The nucleotide fragment identified by the formula above is referred to herein for simplicity as Formula III.

For example, a modified DNA sequence of the invention comprises the Formula III nucleotide sequence and may encode the amino acid sequences identified specifically above, e.g., Fig. 7 [SEQ ID NO:2], Fig. 8 [SEQ ID NO:3]; Fig. 7 [SEQ ID NO:69] and Fig. 8 [SEQ ID NO:70], or other fragments.

In one example of the present invention, the nucleic acid sequence encoding the HA2 subunit protein which contains the native sequence of Formula II has been provided with three silent mutations, which have changed each of the three native arginine-encoding AGG codons each to a preferred arginine codon CGT. These codons encode amino acid numbers 123, 124 and 127 of the H3HA2 subunit protein of the A/Udorn strain identified in Fig. 7. The same codons (and amino acid numbers) are altered in the A/Victoria strain identified in Fig. 8 to provide another example of a modified nucleotide sequence according to this invention.

Thus, with reference to each of Figs. 7 and 8, the native nucleotide sequences encoding the HA2 subunit proteins of the aforementioned viruses [SEQ ID NO:1 and 60], are modified according to this invention at nucleotides 367, 370, and 379. At each of these nucleotide sites, the native A (adenine) is changed to a C (cytosine) and the native nucleotides at sites 369, 372 and 381 in each sequence are changed from a G (guanine) to a T (thymine), resulting in preferred Arg codons.

Other nucleotide sequences encoding the influenza vaccinal polypeptides described herein, or other such influenza proteins or subunits characterized by Formula II may be mutated into novel nucleotide sequences of this invention, i.e., by mutating Formula II into Formula III within those sequences using the first embodiment of the methods of this invention. The silent mutations described herein may be inserted at analogous regions in each nucleotide sequence.

The novel modified H3HA2 nucleotide sequences, whether alone or in association with a nucleotide sequence encoding a fusion partner of a fusion protein of the invention are useful in *E. coli* expression systems. The novel nucleotide sequences of the invention will also encode analogs of the H3HA2 peptides, such as truncated polypeptides (including fragments) and H3HA2 polypeptides, e.g. mutants that retain the epitopes and thus the biological activity of H3HA2. Where the nucleotide sequence encodes a fusion protein, it is anticipated that, because the non-HA2 fusion partner, e.g., NS1 as described below, the fusion peptide provides a means of expressing the protein at high levels and does not appear to play as significant a role in the immunological responses to the HA2 fusion proteins as does the HA2 portion, any number of analogs of this fusion partner can be made.

Typically, the analogs of the nucleotide sequences encoding the HA2 peptides and/or the fusion partner may differ by only 1 to about 4 codon changes, in addition to the nucleotide mutations to the above-identified fragment. Other sequences of this invention include modified nucleotide sequences which encode polypeptides with minor amino acid variations from the natural amino acid sequence of HA2. For example, conservative amino acid replacements may be introduced by altering, deleting or replacing codons of the native sequence, in addition to altering those codons in Formula II according to one embodiment of this method.

Conservative replacements are those that take place within a family of amino acids that are related in their side chains and are well known in the art. For example, it is reasonable to expect that an isolated replacement of a selected amino acid with a conservative replacement of an amino acid with a structurally related amino acid will not have a significant effect on the activity of the protein, especially if the replacement does not involve an amino acid at an epitope of the HA2 polypeptide.

The construction of modified nucleotide sequences and proteins or fusion proteins, given the description herein and conventional methods of protein modification known to one of skill in the art, are believed to be encompassed by this invention.

The novel modified nucleotide sequences of this invention are further characterized by encoding an immunogenic determinant of a modified HA2 subunit of an HA protein, derived from an H3N2 subtype. The encoded protein may contain all or a portion of the H3N2 HA2 sequence, including the Formula I amino acid sequence. The currently preferred embodiment provides a novel DNA sequence encoding an H3HA2 protein or fragment thereof fused in frame to a DNA sequence encoding a portion of the nonstructural influenza protein 1 (NS1). One modified fusion protein-encoding nucleotide sequence is obtained by making mutations according to this invention in the nucleotide sequence encoding the fusion protein $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO:10]. Upon mutation, the nucleotide sequence [SEQ ID NO:58] for this fusion protein [SEQ ID NO:10] is referred to herein as pOTS208NS1H3mut5585.

The modified coding sequences for the HA2 proteins, as well as the coding sequences for NS1 and other viral proteins of influenza virus can be prepared synthetically or can be derived from viral RNA or from available cDNA-containing plasmids by known techniques. For example, see references known to the art which disclose the nucleotide coding sequences for HA from the A/Japan/305/57 strain [Gething *et al.*, Nature, 287:301-306 (1980)]; strain A/NT/60/68 [Sleigh *et al.*, and

Both *et al.*, in Developments in Cell Biology, Elsevier Science Publishing Co., pages 69-79 and 81-89, respectively, (1980)]; strain A/WSN/33 [Davis *et al.*, Gene, 10:205-218 (1980); Hiti *et al.*, Virology, 111:113-124 (1981)]; and fowl plague virus [Porter *et al.* and by Emtage *et al.*, both in Developments in Cell Biology, cited above, at pages 39-49 and 157-168]. Also, influenza viruses, including other strains, subtypes and types, are available from clinical specimens and from public depositories, such as the American Type Culture Collection (ATCC), Rockville, Maryland, U.S.A.

Novel modified nucleotide sequences of this invention may also include allelic variations (naturally-occurring base changes in the species population which may or may not result in an amino acid change) of DNA sequences encoding the H3HA2 protein sequences, and the Formula III fragment [SEQ ID NO:62]. Similarly, DNA sequences having the Formula III fragment, which sequences encode other H3N2 HA2 proteins of the invention include sequences which differ in codon sequence outside of Formula II due to degeneracies of the genetic code or variations in the DNA sequence encoding H3HA2 proteins. Such codon differences may be caused by point mutations or by induced modifications to enhance the activity, half-life or production of the peptide encoded thereby. Also covered by this invention are DNA sequences characterized by the above modification of Formula II into Formula III, which hybridize under stringent conditions with the DNA sequences encoding the HA2 subunit proteins, e.g., H3HA2 proteins, of this invention. DNA sequences which hybridize under non-stringent conditions with the disclosed sequences, but which encode proteins or fragments retaining the biological activities of the H3HA2 proteins, are also included in this invention. Typical conditions for stringent or non-stringent hybridization are known to those of skill in the art [See, e.g., Sambrook *et al.*, cited above].

The actual techniques for producing the mutations described herein are now conventional to the art of genetic engineering, and are readily known and available to one of skill in the art. See, e.g., Sambrook et al, cited above. Such conventional techniques include, for example, site directed mutagenesis, which is available in commercial kits from, e.g. Clonetech and Promega Corporation. Other suitable techniques include, e.g., total gene synthesis and removing the fragment and replacing it with a synthetically generated, mutated fragment. It is anticipated that similar modifications to any H3HA2 sequence having an analogous codon pattern will result in the enhanced expression in *E. coli*, exemplified by the modified H3HA2 sequence.

The mutations described herein are preferentially developed for increased expression of the influenza protein or fusion protein in *E. coli*, which is the preferred host because it can be used to produce the desired proteins safely and cheaply. To circumvent the requirement of ampicillin for plasmid selection in production fermentations, a preferred method of production which uses the modified nucleotide sequences of this invention employs an alternative expression system in which the β-lactamase coding sequence is wholly or partially replaced by a coding sequence for an alternative selectable marker, such as, kanamycin or chloramphenicol.

To aid in expression of the H3HA2 peptides or fusion proteins, these protein sequences or fragments thereof may also be fused to a polypeptide capable of further enhancing expression of these fragments in the selected host system. Ordinarily, such a peptide would contain a leader sequence fragment that provides for secretion of the H3HA2 subunit fragment, in the host cell. The leader sequence fragment typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. There may be processing sites encoded between the leader sequence and the H3HA2 fragment that can be cleaved either *in vivo* or *in vitro*. Alternatively, a promoter sequence may be linked directly with the DNA molecule encoding the H3HA2 fragment. Such polypeptides, promoter and leader sequences are known to those of skill in the art and may be readily selected for expression in the selected host.

Construction of bacterial expression systems, preferably *E. coli* expression systems, including expression vectors and transformed host cells are also within the skill of the art. See, generally, methods described in standard texts, such as Sambrook *et al.*, cited above. The present invention is therefore not limited to any particular vector, nor to any particular purification process from cell lysates or cell medium.

Influenza proteins encoded by the modified nucleotide sequence may be expressed in enhanced manner according to the first embodiment of the method of this invention, or the influenza proteins may be expressed in an enhanced manner by translation from their native sequences by the second embodiment of the method. Additionally, the methods of this invention may be used to enhance the expression of a fusion protein which comprises a protein sequence encoded by the modified nucleotide sequence containing Formula III in place of Formula II in the native nucleotide sequence encoding an HA2 subunit of an HA protein from an H3N2 subtype virus, fused in frame to another protein or protein fragment (a "fusion partner") capable of enhancing expression of the fusion protein.

One of skill in the art may readily select a fusion partner protein or fragment taking into account the desired host cell, i.e., *E. coli*, and utilizing the teachings herein. For the purposes of this invention, the H3HA2 fragment or sequence encoded by a modified nucleotide sequence as described above or the native sequence used in the second embodiment of this method may be fused to any peptide capable of further enhancing its expression in the host cell selected or of increasing its immunogenicity. The method of the present invention does not limit the nature of the "partner" protein or fragment to which the H3HA2 fragment is fused to provide the enhanced expression of the resulting fusion protein.

For example, the influenza protein or fragment bearing the amino acid sequence of Formula I may be fused to a number of conventionally known and used "partner" proteins [See, general texts on expression such as Current Protocols in Molecular Biology, Vol. 2, suppl. 10, publ. John Wiley and Sons, New York, NY, pp. 16.4.1-16.8.1 (1990); Smith *et al*, Gene, 67:31-40 (1988); U. S. Patent No. 4,801,536, among others]. However, it may be desirable that this fusion "partner" protein be an influenza protein sequence or fragment thereof derived from the same or another strain of influenza virus as the HA protein or protein fragment. Preferably, this fusion partner protein is all or a portion of the influenza virus NS1 gene or an HA2 subunit.

In such a fusion protein, a linker sequence may be inserted optionally between the two sequences, i.e., between the sequence encoding the fusion partner and the HA2 protein encoded by the modified nucleotide sequence of this invention or the native sequence for expression according to the second embodiment of the method. This optional linker may provide space between the two protein sequences; and may encode a polypeptide or contain a cleavage site, which is selectively cleavable or digestible by conventional chemical or enzymatic methods. An example of a fusion protein whose expression can be enhanced by a method of this invention is $NS1_{(1-81)}H3HA2_{(1-221)}$ illustrated in Fig. 2 [SEQ ID NO: 10], which comprises the first 81 amino acids of NS1 (derived from an H1N1 subtype virus, A/PR/8/34) fused to the sequences spanning amino acid 1 to 221 of the H3HA2 subunit (amino acids 1-221) via an optional four amino acid linker sequence. Another exemplary fusion protein, $NS1_{(1-81)}H3HA2_{(77-221)}$ SEQ ID NO:72, comprises the first 81 amino acids of NS1 fused to the sequences spanning amino acid 77 to 221 of the truncated H3HA2 subunit. In other embodiments, the NS1 portion may comprise the sequence spanning amino acid residues 1 to amino acids 42 of H1N1. The HA2 fragment may alternatively be fused to a portion of the NS1 peptide derived from a selected Type A virus, e.g., an H3 subtype virus (H3N2).

These proteins, their native nucleotide sequences, and their uses, are described in co-pending U.S. application 07/837,773, filed February 18, 1992, which is incorporated by reference.

As described below in the examples, the host cells used to express these fusion proteins may be modified by the second embodiment of the method of this invention to contain tRNA molecules capable of translating the rare arginine codons of Formula II. See, e.g., Example 25. Alternatively, the nucleic acid sequence encoding these and other suitable H3HA2 proteins or H3HA2-containing proteins, i.e. those comprising a native Formula II sequence [SEQ ID NO:9], may be modified by the first embodiment of the method of this invention to replace Formula II with the Formula III sequence to increase the expression of the encoded protein in *E. coli* according to the method of this invention.

The proteins and fusion proteins whose expression is enhanced by the methods of this invention may be employed in vaccine compositions. Several of the specific influenza proteins or fusion proteins described herein, which are produced according to the methods of this invention, have demonstrated the ability to stimulate or produce a protective immune response capable of recognizing an influenza virus or influenza virus-infected cells and protecting the vaccinated mammal against disease caused thereby. This protective response is desirably a T cell response, produced in the substantial absence of vaccine-induced neutralizing antibody. Such H3HA2 proteins and fusion proteins are capable of inducing T helper cells, particularly cytotoxic T lymphocytes, in the absence of neutralizing antibodies.

Pharmaceutical vaccine compositions can contain an effective immunogenic amount of a selected H3HA2 protein produced according to this invention or encoded by a modified nucleotide sequence of this invention in admixture with a suitable adjuvant in a nontoxic and sterile pharmaceutically acceptable carrier. Suitable carriers for vaccine use, as well as other vaccine formulation additives and adjuvants, are well known to those of skill in the art. See, e.g., European Patent Application No. 366, 238, published May 2, 1990; and European Patent Application No. 366,239, published May 2, 1990. Such compositions may be effectively administered to human and animal patients to induce the appropriate immune response. The details of dosage and treatment using such compositions are also described in the above-cited published patent applications.

The following examples illustrate methods for preparing H3HA2 and BHA2 fusion proteins of the invention and demonstrate the subtype specific

protection against heterologous virus induced upon vaccination with the H3HA2 proteins. The following examples also illustrate methods for preparing the modified DNA sequences of the invention. All of these examples are illustrative only and do not limit the scope of the invention.

EXAMPLE 1 - PLASMID pMS3H3HA

Plasmid pFV88 contains the entire 221 amino acid length HA from A/Udorn, an H3 subtype virus [C. J. Lai et al, Proc. Natl. Acad. Sci. USA, 77:210-214 (1980)], which HA nucleic acid sequence is illustrated in Fig. 1 [SEQ ID NO: 1]. This plasmid was cut with Pst I. The resulting 1900 bp fragment, which contains the entire HA (HA1 and HA2) fragment and some GC tailing, was then inserted into pUC18 [Bethesda Research Laboratories]. The resulting plasmid is termed pMS3 or pMS3H3HA.

EXAMPLE 2 - pMG1

Plasmid pAPR801 is a pBR322-derived cloning vector which carries the NS1 coding region (A/PR/8/34). It is described by Young et al, in The Origin of Pandemic Influenza Viruses, ed. by W. G. Laver, Elsevier Science Publishing Co. (1983).

Plasmid pAS1 is a pBR322-derived expression vector which contains the P_L promoter, an N utilization site (to relieve transcriptional polarity effects in the presence of N protein), and the cII ribosome binding site including the cII translation initiation codon followed immediately by a BamHI site. It is described by Rosenberg et al, in Methods Enzymol., 101:123-138 (1983).

Plasmid pAS1ΔEH was prepared by deleting a non-essential EcoRI-HindIII region of pBR322 origin from pAS1. A 1236 base pair BamHI fragment of pAPR801, containing the NS1 coding region in 861 base pairs of viral origin and 375 base pairs of pBR322 origin, was inserted into the BamHI site of pAS1ΔEH. The resulting plasmid, pAS1ΔEH/801, expresses authentic NS1 (230 amino acids). The plasmid has an NcoI site between the codons for amino acids 81 and 82 and an NruI site 3' to the NS sequences. The BamHI site between amino acids 1 and 2 is retained.

Plasmid pMG27N, a pAS1 derivative [Mol. Cell. Biol., 5:1015-1024 (1985)], was cut with BamHI and SacI and ligated to a BamHI/NcoI fragment encoding the first 81 amino acids of NS1 from pAS1ΔEH801 and a synthetic DNA NcoI/SacI fragment of the following sequence:

SEQ ID NO: 35:
5'-CATGGATCATATGTTAACAGATATCAAGGCCTGACTGACTGAGAGCT-3'
SEQ ID NO: 36:
3'- CTAGTATACAATTGTCTATAGTTCCGGACTGACTGACTC -5'.

The resulting plasmid, pMG1, allows the insertion of DNA fragments after the first 81 amino acids of NS1 in any of the three reading frames within the synthetic linker fragment followed by termination codons in all three reading frames.

EXAMPLE 3 - pMG1H3HA

Plasmid pMG1, described above in Example 2, was digested with NcoI and XbaI, releasing a 54 bp fragment, which was discarded. Plasmid pMS3H3HA, described in Example 1 above, was digested with HhaI and XbaI, and a 701 bp fragment containing the coding sequence for the HA2 subunit of influenza strain A/Udorn (H3N2) was isolated, as illustrated in Fig. 1 [SEQ ID NO: 1].

Synthetic oligonucleotides were annealed to generate an NcoI 5' overhang sequence (at the 5' end) and a HhaI 3' overhang sequence (at the 3' end). The sequence of these oligonucleotides is as follows:

SEQ ID NO: 37: 5'-CATGGGCGCCCATATGGGCATATTCGGCG-3'
SEQ ID NO: 38: 3'- CCGCGGGTATACCCGTATAAGCC-5'.

The annealing reaction was performed as follows. The annealing mixture was made up of 2.5μL each of 5' oligo (1.3 μg/μL), the 3' oligo (1.2 μg/μL), and added water (15 μL) to a final volume of 20 μL. The reaction tubes were then placed in 4 mL culture tubes containing water which had been heated to 65°C for 10 minutes and allowed to cool down slowly. The tubes were then put on ice and used immediately for ligation.

This three part ligation generates pMG1H3HA2$_{(1-221)}$ [SEQ ID NO: 9] which codes for the first 81 amino acids of NS1 fused to four amino acids donated from the linker and amino acids 1-221 of the HA2 subunit. This sequence is illustrated in Fig. 2 [SEQ ID NO: 9 & 10]. This molecule is also designated NS1$_{(1-81)}$H3HA2$_{(1-221)}$ [SEQ ID NO: 9 & 10] or H3C13.

EXAMPLE 4 - NS1$_{(1-81)}$H3HA2$_{(77-221)}$ [SEQ ID NO: 11 & 12]

pMS3H3HA, described in Example 1 above, was digested with EcoRI and end-filled (Klenow). Subsequently, the vector was digested with XbaI. A 487 bp fragment, which contains the coding sequence for amino acids 77-221 of

the HA2 subunit, was isolated and ligated to the HpaI and XbaI sites of pMG1. The resulting vector codes for a fusion polypeptide containing amino acids 1-81 of NS1 fused to amino acids 77-221 of the HA2 subunit. This molecule has been termed $NS1_{(1-81)}H3HA2_{(77-221)}$ and is illustrated in Fig. 3 [SEQ ID NO: 11 & 12].

EXAMPLE 5 - $pMG_{42}BLHA2$

To derive a vector similar to pMG1 (described in Example 2), which contains the coding region for the first 42 amino acids of NS1 rather than the first 81 amino acids of NS1, pMG1 was digested with BamHI and NcoI and ligated to the BamHI/NcoI fragment encoding amino acids 2 to 42 of NS1 from $pNS1_{42}TGF$ α. $pNS1_{42}TGF\alpha$ is derived when pAS1ΔEH801 is cut with NcoI and SalI and ligated to a synthetic DNA encoding human TGFα as an NcoI/SalI fragment. $pNS1_{42}TGF\alpha$ encodes a protein comprised of the first 42 amino acids of NS1 and the mature TGFα sequence. The NS1 portion of $pNS1_{42}TGF\alpha$ contains an amino acid change from Cys to Ser at amino acid 13.

The resulting plasmid, termed $pMG_{42}A$, was then modified to contain an alternative synthetic linker after the $NS1_{42}$ sequence with a different set of restriction enzyme sites within which to insert foreign DNA fragments into the three reading frames after the $NS1_{42}$. This linker has the following sequence:

SEQ ID NO: 39:

5'-

CATGGATCATATGTTAACAAGTACTCGATATCAATGAGTGACTGAAGCT-

3'

SEQ ID NO: 40:

3'- CTAGTATACAATTGTTCATGAGCTATAGTTACTCACTGACT -5'.

The resulting plasmid is called $pMG_{42}B$. This vector is needed to contain the neomycin phosphotransferase-1 (NPT-1) gene which confers kanamycin resistance.

As described in Shatzman and Rosenberg, Met. Enzymol., 152:661-673 (1987), pOTS207 is a pAS derived cloning vector which carries the kanamycin resistance gene from Tn903 [Berg et al, Microbiology, ed. D. Schlessinger, pp. 13-15, American Society for Microbiology (Washington, DC 1978); Nomura et al, The Single-Stranded DNA Phages, ed. D. Denhardt et al, pp.467-472, Cold Spring Harbor Laboratory (New York 1978); Castellazzi et al, Molecul. Gen. Genet., 117:211-218 (1982)]. It was constructed by digesting plasmid pUC8 [Yanisch-Perron et al, Gene, 33:103-119 (1985)], with BamHI and ligated to a BclI fragment containing the kanamycin gene from Tn903. The resulting plasmid, pUC8-Kan, was digested with EcoRI and PstI, and the fragment containing the kanamycin gene

was inserted between the EcoRI and PstI sites of pOTSV [Shatzman and Rosenberg, cited above]. The resulting plasmid is pOTS207.

The pOTS207 was digested with EcoRI and PstI, and the 1467 bp fragment containing the kanamycin resistance gene was isolated. Synthetic oligonucleotides:

SEQ ID NO: 41: 5' AATTCGTACCTA 3'

SEQ ID NO: 42: 3' GCATGGATCTAG 5'

were made to link the NPT-1 gene to pMG42B vector. $pMG_{42}B$ was digested with BglII and PstI. The EcoRI/PstI NPT-1 gene fragment and the synthetic oligo linker were ligated to the digested $pMG_{42}B$. The resulting plasmid, $pMG_{42}Kn$ allows fusions, in three different reading frames, to the $NS_{1\text{-}42}$ gene, while allowing antibiotic selection with kanamycin.

Plasmid pBHA is a pBR322-derived vector, containing the complete nucleotide sequence of the HA gene of a Type B influenza virus (B/Lee/40). It is described by Krystal et al, Proc. Natl. Acad. Sci. USA, 79:4900-4804 (1982). pBHA was digested with RsaI and a 813 bp fragment containing the HA subunit was isolated. This fragment was ligated into plasmid $pMG_{42}Kn$ (described above) that had been digested with ScaI. During the cloning, a nucleotide base (T) was deleted from the ScaI recognition site shifting the gene out of the reading frame. The vector was digested with NcoI, and filled-in using Klenow, putting the gene back into the reading frame.

The resulting construct, $pMG_{42}BLHA2$ [SEQ ID NO: 14], expresses a fusion polypeptide containing amino acids 1-42 of NS1 and 41-233 of the HA2 subunit. This construct contains the Cys to Ser change at amino acid 13 of the NS1 portion of the fusion peptide.

In preliminary studies with this construct, vaccinated laboratory mice demonstrated protection from challenge with Type B influenza in the absence of neutralizing antibody for the virus.

EXAMPLE 6 - PREPARING SEED VIRUS AND RAISING ANTISERA

The seed virus, A/Udorn, was prepared according to the procedures described in P. Palese and J. Schulman, Virol., 57:227-237 (1974). Briefly, this technique is as follows.

Influenza virus strain A/Udorn was inoculated in 10-day old embryonated hen's eggs into the allantoic cavity. The eggs were incubated for 24-48 hours at 35°C then chilled at 4°C overnight. A portion of the eggshell over the airsac was removed and the allantoic fluid was aseptically removed using a

10-ml syringe. The fluid was centrifuged at low speed (3,000 x g) to remove particulates. This clarified supernatant was centrifuged at high speed using an SW28 Beckman rotor at 27,000 rpm (4°C for 90 minutes), resulting in the virus pellet. The virus was resuspended in 10 mM Tris (pH 7.5) containing 100 mM NaCl, 1 mM EDTA and repelleted as before. The virus was layered on 30-60% sucrose gradient in 1 mM EDTA (NTE) and spun for 3-5 hours at 25,000 rpm. The band in the middle of the tube was withdrawn, diluted in NTE and centrifuged at 27,000 rpm for 90 minutes. The pellet was suspended in phosphate-buffered saline (PBS). These viral particles were used as immunogens for preparation of antisera.

Antisera was prepared as follows. 100-200 micrograms of purified virus in complete Freund's adjuvant was injected into the subscapula of a New Zealand White rabbit. A second injection in incomplete Freund's adjuvant was done 4 weeks later, and the animals were bled and antisera collected 7-10 days later.

EXAMPLE 7 - EXPRESSION OF H3HA2 FUSION PROTEINS

A. $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 9 & 10]

The plasmid $pMG1H3HA2_{(1-221)}$ [SEQ ID NO: 9] was transfected into E. coli strain AR58 [SmithKline Beecham Pharmaceuticals]. Cultures were grown at 32°C to mid-log phase at which time cultures were shifted to 39.5°C for 2 hours. The E. coli cell pellets containing the recombinant polypeptide were then stored at -70°C until used.

Production of the $NS1_{(1-81)}H3HA2_{(1-221)}$ protein [SEQ ID NO: 10] was confirmed by Western blot analysis [Towbin et al, Proc. Natl. Acad. Sci. U.S.A., 76:4350 (1979)] using antisera prepared against A/Udorn virus, as described in Example 5. A major immunoreactive species was found at a molecular weight of 35,050 daltons.

B. $NS1_{(1-81)}H3HA2_{(77-221)}$ [SEQ ID NO: 11 & 12]

The plasmid encoding the $NS1_{(1-81)}H3HA2_{(77-221)}$ peptide [SEQ ID NO: 12] was expressed as described in part A above. Production of this peptide was confirmed by Western blot analysis, as described above. A major immunoreactive species was found at a molecular weight of 26,697 daltons.

EXAMPLE 8 - PARTIAL PURIFICATION OF H3HA2 FUSION PROTEINS

E. coli cell pellets containing the recombinant polypeptides, prepared as described in Example 6, were stored at -70°C until used. E. coli cells were thawed and resuspended in lysis buffer A (50 mM Tris-HCl, 5% glycerol, 2 mM EDTA and 0.1 mM DTT, pH 8.0) at 10 mL/gram. The stirred suspension was then

treated with lysozyme (0.2 mg/mL) for 45 minutes at room temperature and sonicated 2x for 2-3 minutes each time by a Sonicator. The resultant suspension was treated with 0.1% DOC for 60 minutes at 4°C, then centrifuged at 25,000 x g. The pellet was resuspended by sonication in 50 mM glycine pH 10.0, 5% glycerol, 2 mM EDTA and then the suspension was treated with 1% Triton X-100 [J.T. Baker Chemicals Co.] at 4°C for 60 minutes and centrifuged as above.

The resulting pellet was solubilized in 50 mM Tris, 8 M urea, pH 8.0 and centrifuged to remove any insoluble material. This solubilized material is dialyzed against 10 mM Tris, 1 mM EDTA, pH 8.0 followed, again, by centrifugation of insoluble material. The solubilized material is designated as "crude" material and is used in in vitro and in vivo mouse assays. At this point, the material is approximately 40 - 50% pure.

The "crude" material was electrophoresed through an SDS-PAGE and the appropriate H3HA2 protein bands were visualized by KCl staining according to D. Hager et al, Anal. Biochem, 109:76-86 (1980). The band was cut-out and eluted electrophoretically by the "S&S Elutrap Electro-Separation System" [Schleicher & Schuell]. The electro-eluting buffer was the Tris-glycine. A concentrated and eluted sample was obtained and exhaustively dialyzed against 0.01 M NH_4HCO_3 and 0.02% SDS [M. Hunkapiller et al, Method. Enzymol., 91:227-236 (1983)]. This sample was frozen quickly by dry ice and lyophilized to complete dryness. The lyophilized material was brought back into solution using 50 mM Tris pH 8.0 and used for in vitro and in vivo mouse assays.

Following this gel elution step, the protein is usually greater than 75% pure.

EXAMPLE 9 - CONSTRUCTION OF POTS208 VECTORS

pOTSV is described in Devara et al, Cell, 36: 43-49 (1984). Briefly, this vector is a pAS1 derivative with t-oop inserted at the NruI site and a synthetic oligonucleotide encoding SacI, XhoI and XbaI restriction sites inserted at the SalI site (which is destroyed).

A. pOTS208

pOTS208 was prepared by digesting pOTSV with EcoRI and ScaI, followed by fill in reaction using Klenow. Tn5 Plasmid DNA [described in R. Jorgensen et al., Mol. Gen. Genet., 177:65-72 (1979)] was digested with HindIII and SmaI, followed by a fill in reaction using Klenow yielding a 1323 bp fragment encoding for neomycin phosphotransferase-2 gene (NPT-2). This fragment is described in detail in Rothstein et al., Cell, 19:795-805 (1980) and Jorgensen, cited

above. This fragment and the above digested vector were ligated together to create pOTS208, which is kanamycin resistant.

B. pOTS208H3C13

pMG1H3HA2(1-221) (Example 3) was digested with BamHI and XbaI, releasing two fragments: an 806 bp BamHI fragment and a 160 bp BamHI/XbaI fragment. These fragments together code for $NS1_{(1-81)}H3HA2_{(1-221)}$. A three part ligation between the two fragments and BamHI/XbaI digested pOTS208 (part A) yield pOTS208H3C13 which utilizes NPT-2 for kanamycin resistance.

C. pOTS208NS181Nco

pOTS208H3C13 (part B) was digested with BglII. pSelect [Promega] was cut with BamHI and ligated with the BglII fragment, resulting in pSelectNPTII102. Transformation into *E. coli* JM101 [ATCC *E. coli* 33876] was followed by selection on kanamycin and tetracycline plates. KanR was conferred by the NPT2 region from pOTS208H3C13. Some lambda sequence was also on the BglII fragment. Oligo 4852, SEQ ID NO: 49 GCATCGCCATGAGTCACGACG, was used to mutate the NcoI site to CCATGA in pSelectNPTII102, resulting in pSelectNPTII102-8. This vector was cut with BstEII and BssHII. pOTS208H3C13 was cut with BstEII, BssHII and SphI, and fragment exchange generated pOTS208NS181H3HA2-26. This clone has the NcoI site of NPT2 mutated. pOTS208NS181NS181H3HA2-26 was cut with NcoI and SalI, filled in and ligated with Linker 1041 [New England Biolabs] to insert a KpnI site and regenerate the NcoI site. This step also deletes the H3C13 region. The unique XbaI site of the parent pOTS208 vector is downstream of the deletion. The resulting vector is pOTS208NS181Nco.

EXAMPLE 10 - MODIFICATION OF GENE ENCODING H3HA2 FUSION PROTEIN

In order to increase yield of the H3HA2 protein, silent mutations to certain rare arginine codons were made to the coding sequence of the H3HA2 protein. These nucleotide changes resulted in no change in the protein sequence.

A mutant H3C13 protein was prepared by mutating the nucleotide sequences of the fusion protein prepared according to Example 3 above. Site directed mutagenesis using the Altered Sites System [Promega Corporation] according to the manufacturer's directions was used to change nucleotide numbers, 622, 625, and 634 (A to C) and 624, 627, and 636 (G to T) of nucleotide sequences [SEQ ID NO:9] encoding the $NS1_{(1-81)}H3HA2_{(1-221)}$ fusion protein of Fig. 2 [SEQ ID NO:10], thereby changing the codons at these regions from AGG to CGT,

both encoding Arg. These changes correspond to nucleotide numbers 367, 370, and 379 (A to C) and 369, 372, and 381 (G to T) of the HA2 fragment of Fig. 2 [SEQ ID NO: 58].

Fig. 2 illustrates the modified nucleotide sequences of the fusion protein [SEQ ID NO: 10] by contrast with the nucleotide sequence [SEQ ID NO: 9] of the "unmodified" fusion gene (nucleotide changes above sequences of unmodified gene). Mutagenesis on this sequence was carried out according to the method provided with the pSelect kit from Promega.

A. $\underline{NS1_{(1-81)}H3HA2_{(1-221)}}$ [SEQ ID NO: 10]

Briefly, cloning for the mutagenesis was performed as follows. The pSelect plasmid [Promega] and pMG1H3HA2 (Example 3) were each digested with HindIII. These two plasmids were ligated together and selected on tetracycline plates. The resulting vector is pSelH3HA2. Mutagenesis was performed according to Promega's kit. The following oligonucleotide was used: SEQ ID NO: 43: 5'-AAACTGTTTG AAAAAACACG TCGTCAACTG CGTGAAAATG CTGACGACAT GGGC -3'.

Clones were verified by restriction endonuclease HincII. The resulting plasmid, pSelH3HA2mut5585 was digested with NcoI and XbaI, and a 748 bp fragment coding for the H3HA2mut5585 polypeptide was isolated.

pOTS208NS181Nco (Example 9C) was digested with NcoI and XbaI. The ligation of linear pOTS208NS181Nco and the 748 bp fragment resulted in pOTS208NS1H3mut5585 [SEQ ID NO:7]. This vector codes for the polypeptide, $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO:10].

B. Expression of mutated gene encoding H3C13 protein

The plasmid of A was transfected into *E. coli* strain AR58 [SmithKline Beecham]. Cultures are grown at 32°C to mid-log phase at which time cultures are shifted to 39.5°C for two hours. The *E. coli* cell pellets containing the recombinant polypeptide are then stored at -70°C until used. Production of the $NS1_{(1-81)}H3HA2_{(1-221)}$ protein [SEQ ID NO:10] is confirmed by Western blot analysis [Towbin *et al.*, Proc. Natl. Acad. Sci. U.S.A., 76:4350 (1979)] using antisera prepared against A/Udorn virus, as described in Example 4. A major immunoreactive species is expected at a molecular weight of approximately 35,00 daltons.

The expression levels obtained are about 50-100% higher than those obtained by the expression of the unmodified coding sequences in the same expression system.

C. Construction of Alternative H3 Mutant

pSelH3HA2mut5585 (part A) was subjected to site-directed mutagenesis, as described above. Oligo SEQ ID NO: 44 TGTGACAATGCTTGCATCGGTTCAATCCGTAATGGTACTTATGACCA TGATG, was used and clones were verified by restriction endonuclease Rsa1. The resulting plasmid, pSelH3HA2mut2 was digested with Nco1 and Xba1, and an approximately 748 bp fragment encoding for the H3HA2mut2 polypeptide was isolated. pOTS208NS181Nco was digested with Nco1 and Xba1. The ligation of linear pOTS208NS181Nco (Example 9C) and the 748bp fragment resulted in pOTS208NS1H3mut2. This vector codes for the $NS1_{(1-81)}H3HA2_{(1-221)}$ polypeptide [SEQ ID NO: 10].

Example 11 - PLASMID pD

Plasmid pAS1_EH/801 (described above in Example 2) was cut with BglII, end-filled with DNA polymerase I (DNApolI; Klenow), and ligated closed, thus eliminating the BglII site. The resulting plasmid $pBgl^-$ was digested with NcoI, end-filled with DNApolI (Klenow), and ligated to a BglII linker. The resulting plasmid, pB4, contains a BglII site within the NS1 coding region. Plasmid pB4 was digested with BglII and ligated to a synthetic DNA linker of the sequence:

SEQ ID NO: 45: 5'-GATCCCGGGTGACTGACTGA -3'
SEQ ID NO: 46: 3'- GGCCCACTGACTGACTCTAG-5'.

The resulting plasmid, pB4+, permits insertion of DNA fragments within the linker following the coding region for first 81 amino acids of NS1 followed by termination codons in all three reading frames. Plasmid pB4+ was digested with XmaI (cuts within linker), end-filled (Klenow), and ligated to a 520 base pair PvuII/HindIII, end-filled fragment derived from the HA2 coding region. The resulting plasmid, pD, codes for a protein [SEQ ID NO: 18] comprised of the first 81 amino acids of NS1, three amino acids derived from the synthetic DNA linker (Gln-Ile-Pro), followed by amino acids 65-222 of the HA2.

Expression is obtained by transfecting pD into a desired *E. coli* strain, preferably LW14, using standard techniques. Purification may be by standard techniques or, preferably, as described in Example 18 below.

EXAMPLE 12 - H3 SUBTYPE HETEROLOGOUS PROTECTION ELICITED BY VACCINATION WITH $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10]

Mice (NIH/Swiss; 15 per group) were vaccinated subcutaneously with 50 or 10 µg $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 9 & 10] in aluminum

hydroxide on days 0 and 21. The mice were boosted intraperitoneally on day 42 with the protein without adjuvant. On day 47, mice were challenged intranasally with 2 - 3 LD_{50} doses of either A/PR/8/34 (H1N1) or A/HK/68 (H3N2) virus, and survival was monitored through day 21. This represents a heterologous challenge (A/PR/8/34) and an H3 heterosubtypic challenge, since the $NS1_{(1-81)}H3HA2_{(1-221)}$ construct [SEQ ID NO: 9 & 10] was derived from A/Udorn/72 cDNA. The control group received adjuvant (CFA) only.

The results in Table 1 below show that survival in mice vaccinated with $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10] and challenged with A/HK/68 (80-93%) was significantly higher than in control mice which were injected with adjuvant only (26% survival). In contrast, vaccination with $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10] did not confer protection against challenge with A/PR/8/34, an H1N1 strain (0-26% survival). Thus protection elicited by $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10] is selective for antigenically diverse virus strains within the H3 subtype.

Likewise, vaccination with the D protein ($NS1_{(1-81)}HA2_{(65-222)}$ [SEQ ID NO: 18], derived from the H1N1 subtype) elicits protection from heterosubtypic challenge with H1N1, but not the H3N2 subtype [S. Dillon et al, Nature, (1992); Mbawuike et al, Faseb. J., 5:A1362 (abs. 5749 and Table 1)]. These results in outbred mice also suggest that the response to the H1 and H3 proteins will not be restricted to a limited number of individuals with certain major histocompatibility alleles, and therefore the vaccine will be effective in a majority of individuals.

Table 1

Percent Survival After Challenge:

Immunization	HA Subtype	A/PR/8/34 (H1N1)	A/HK/68 (H3N2)
50 µg $NS1_{(1-81)}H3HA2_{(1-221)}$	H3	26	80*
10 µg $NS1_{(1-81)}H3HA2_{(1-221)}$	H3	0	93*
10 µg $NS1_{(1-81)}HA2_{(66-222)}$	H1	67*	13

A/HK/68 virus	H3	60*	100*
Control (Al^{+3})	-	0	26

$p \leq 0.05$ vs. control in Fishers exact probability test

Vaccination of mice with live homologous (A/HK/68) virus provided complete or partial protection, reflecting protection mediated by neutralizing antibody (homologous H3N2 challenge) and/or CTL (heterologous H1N1 challenge), respectively.

Duration of protective immunity was tested by immunizing mice subcutaneously with the recombinant influenza protein plus adjuvant on days 0 and 21. Some mice were also given an ip injection of the protein (without adjuvant) on day 42. Mice were challenged with A/HK/68 (H3N2) on day 47, four weeks after the second injection. Control mice were immunized as described above for Table 1, where an ip injection was given at week 6 (5 days prior to challenge). The results in Table 2 show that CB6F$_1$ mice (15 per group) were significantly protected when challenged with the A/HK/68 heterologous H3 virus strain 5-28 days after the last injection.

Table 2

Dose (μg per injection) of NS1$_{(1-81)}$H3HA2$_{(1-221)}$	Adjuvant	Injection Schedule	Percent Survival
50 μg	CFA	0,21	86*
50 μg	CFA	0,21,42	100*
0 μg	CFA	0,21	6
50 μg	Al^{+3}	0,21	93*
50 μg	Al^{+3}	0,21,42	93*
0 μg	Al^{+3}	0,21	0

*$p \leq 0.05$ v. control in Fisher's exact probability test

EXAMPLE 13 - TYPE A CROSS-PROTECTION WITH D AND H3C13 PROTEIN

Mice (CB6F$_1$) were divided randomly into six groups, with fifteen in each group. The mice were injected subcutaneously with proteins in Al^{+3} (100 μg) on days 0 and 21, and then were challenged with 2-3 LD$_{50}$ doses of virus on day 49. Survival was monitored through day 21. The results of this study are illustrated in Table 3 below. For convenience, NS1$_{(1-81)}$H3HA2$_{(1-221)}$ [SEQ ID NO: 10] is referred to as H3C13 in the table below.

Table 3

Percent Survival After Challenge with:

	Immunization	HA Subtype	A/PR/8/34 (H1N1)	A/HK/68 (H3N2)
1.	50 μg H3C13 50 μg D	H3 H1	73*	73*
2.	10 μg H3C13 10 μg D	H3 H1	67*	100*
3.	1 μg H3C13 1 μg D	H3 H1	86*	73*
4.	50 μg H3C13	H3	7	73*
5.	50 μg D	H1	47**	7
6.	Al^{+3} control	-	7	0

* $p \leq 0.001$ vs. control group

** $p \leq 0.03$ vs. control group

This data demonstrates that mice immunized with a mixture of the D protein and H3C13 protein in aluminum adjuvant were protected against challenge with either A/PR/8/34 (H1) or A/HK/68 (H3) virus. In contrast, mice immunized with the D protein were protected against H1 but not H3 challenge. Likewise, mice immunized with the H3C13 protein were protected against the H3 but not the H1 challenge. Therefore, the combination of the D protein and the H3C13 proteins elicited protection against the currently circulating subtypes of influenza A virus. Thus, this combination represents a subtype cross-protective vaccine.

EXAMPLE 14 - CREATION OF pEA181KNRBS3 VECTOR

pMG1 (Example 2) and pMG42Kn (Example 5) were both digested with BamHI and NcoI. A 236 BamHI/NcoI fragment containing the coding sequence for amino acid sequence spanning residues 1 to 81 of the NS1 gene was isolated from pMG1. The digested pMG42Kn and the 236 bp fragment were ligated

together and transformants were selected on LB and kanamycin agar plates. The resulting vector pMG181Kn(cII) maintains all regulatory elements of pMG42Kn except the NS1 (aa1-42) sequence is replaced with the NS1 (aa1-81) sequence.

pMG181Kn(cII) described above was digested with BstXI and BamHI. The following linker encoding ribosome binding site (RBS3) is cloned in the digested vector, replacing the cII RBS. The linker sequence is:

5' TAAGGAGGATATAACATATG [SEQ ID NO: 47]
3' <u>TGGA</u>ATTCCTCCTATATTGTATAC<u>CTAG</u> 5' [SEQ ID NO: 48].
BstXI BamHI

The resulting vector is pMG181KnRBS3.

To generate pEA181KnRBS3, a 1.2 kb EcoRI/BglII fragment from similarly digested pOTSV containing the lambda rexArexB region was cloned into mp18 [Gibco/Bethesda Research Labs] and mutagenized to create silent mutations in the two NdeI sites in this region. The mutations were CATATG to CATGTG in both sites. One site is in the rexA and the other in the rexB. The mutagenized fragment was inserted into both EcoRI/BglII digested pMG181Kn(cII) and similarly digested pMG181KnRBS3, resulting in the plasmids pEA181Kn(cII) and pEA181KnRBS3, respectively. pEA181KnRBS3 has the useful properties of the pMG vectors, plus the additional attribute of nalidixic acid induction.

An EcoRI/PstI fragment containing the ampR gene of pBR322 was then inserted into EcoRI/PstI digested pEA181Kn(cII) and pEA181KnRBS3 to create pEA181CIIamp and pEA181RBS3amp, respectively. These plasmids are rexB+ and should be nalidixic acid inducible, in contrast to pMG1 and its descendants, which are rexB- and cannot be induced with nalidixic acid. The mutant EcoRI/BglII region was functionally examined by cloning it into a pMG1 vector carrying galK and demonstrating induction of galK with nalidixic acid.

<u>EXAMPLE 15 - CREATION OF VECTOR FOR PRODUCTION OF NS1$_{(1-81)}$BHA2$_{(1-223)}$</u>

Plasmid pOTS208BLeeHA2 was created as follows. An EcoRI fragment encoding the B/Lee HA region from plasmid pBHA (Example 5) was cloned into pSelect to generate pSelectPBHAS2. Site-directed mutagenesis inserted an NcoI site at the start of HA2, resulting in an N-terminus: MET GLY PHE PHE, and a C terminus of SER ILE CYS LEU. The resulting construct is called pSelectPBHAS2-B1. This plasmid was cut with NcoI and XbaI (a site in the polylinker of pSelect downstream of the HA gene), and ligated into NcoI/XbaI digested pEA181KnRBS3, described above, to generate pEA181BLeeB1-1. A

BamHI/EcoRI, filled in, fragment was cut out of pEA181BLeeB1-1 and ligated into pOTS208 (Example 9A), that had been digested with XbaI, filled in, and BamHI. The EcoRI and XbaI sites were regenerated by the ligation. This extra cloning step was necessary because there was no convenient cloning site to fuse the gene to $NS_{(1-81)}$ in pOTS208. pEA181KnRBS3 (described above) and pOTS208 have unique BamHI and XbaI sites to facilitate fragment exchanges.

A BamHI/XbaI fragment of about 1011 bp encoding the $NS_{(1-81)}BLHA2_{(1-223)}$ sequence from plasmid pOTS208BLeeHA2 was isolated and ligated into vector pSelect-1 [Promega], which was also digested with BamHI and XbaI. The resulting construct is called pSelBC13. This vector contains the coding sequence for $NS1_{(1-81)}BHA2_{(1-223)}$, also termed BC13 [SEQ ID NO: 57].

EXAMPLE 16 - CREATION OF VECTOR FOR PRODUCTION OF BC13mut2

Mutagenesis was carried out on the pSelBC13 using Promega's protocol and oligonucleotide 5492, SEQ ID NO: 50 GGAGGATGGGAAGGACTCATTGCAGGTTGG. This mutagenesis changed the ATG codon within the HA2 portion of the molecule to CTC (MET to LEU). The resulting plasmid is called pSelBC13mut5492. This plasmid was then digested with NcoI and XbaI, releasing a digestion fragment encoding for HA2, and ligated into pOTS208NS181Nco (Example 9C) that had been digested with NcoI and XbaI. The resulting construct, pOTS208NS1BLHA2mut5492 codes for the same polypeptide of pOTS208BLeeHA2, (i.e. BC13), except the internal translation start is eliminated at amino acid position 98 of the fusion protein. This protein is $NS1_{(1-81)}BHA2_{(1-223)}$(met-leu) [SEQ ID NO: 55].

A HindIII fragment of approximately 1 kb encoding NS1 (amino acid residues 7-81) and BLee HA2 (amino acid residues 1-223) and which contained the MET to LEU changes from plasmid pOTS208NS1BLHA2mut5492 was isolated. This fragment was ligated into the HindIII site of vector pSelect-1, resulting in pSelBC13mut5492. Mutagenesis was carried out using Promega's protocol and the following oligos 5920, 5921 and 5939, respectively:

SEQ ID NO: 51

CTCTGCTGTAGAAATCGGTAACGGTTGCTTTGAAACCAAAC

SEQ ID NO: 52

GGTTTCTTGGAAGGTGGTTGGGAAGGTCTCATTGCAGGTTGGCACGG

SEQ ID NO: 53

GCTTTCCAACGAAGGTATCAATCAACAGTGAAGACGAGCATCTCTTGG.

This mutagenesis created the following silent codon changes in the HA2 region:

The codons for GLY at positions 93, 94, 97, 187, 215, and 217 were each mutated from GGG to GGT; the codons for ILE at positions 188, 189, and 214 were each changed from ATA to ATC; the codon for ASP at position 193 was changed from GAT to GAC; and the codon for ASN at position 216 was changed from AAT to AAC.

The resulting plasmid was called pSelBC13mut2. This plasmid was then digested with NcoI and XbaI, releasing a fragment of about 775 bp encoding for HA2. This fragment was ligated into pOTS208NS181Nco (described above), that had been digested with NcoI and XbaI. The resulting construct, pOTS208NS1BLmut2 (see Fig. 5 [SEQ ID NO: 54]), codes for the same polypeptide [SEQ ID NO: 55] as pOTS208NS1BLHA2mut5492, except for the silent codon changes.

EXAMPLE 17 - EXPRESSION OF FUSION PROTEIN

pOTS208NS1BLmut2 [SEQ ID NO: 54] is transfected into a suitable host cell, preferably an *E. coli* strain and expressed essentially as described for the H3 proteins described above. Strain LW14 is a derivative of *E. coli* K-12 strain W3110 [ATCC *E. coli* 27325]. The transducing phage P1 [*E. coli* ATCC 25404-B1] was grown on *E. coli* K-12 strain AR58, described above, the genotype of which is thr-galE::Tn10 λCI857 bio-uvrB- rpsL. Phenotypically, strain AR58 requires threonine, biotin for growth, is sensitive to UV light and DNA damaging agents, cannot use galactose as a carbon source, and is resistant to streptomycin. Strain W3110, a prototroph, is incubated with the phage and plated onto a medium containing tetracycline to select for the transduction of the Tn10 element. The P1 phage picks up the segment of DNA containing the Tn10 and brings with it the λ CI857 bio- uvrB-. Thus the strain LW14 has the following genotype: galE::Tn10λ CI857 bio- uvrB-. Phenotypically, strain LW14 requires biotin for growth, is sensitive to UV light and DNA damaging agents, and cannot use galactose as a carbon source.

EXAMPLE 18 - PURIFICATION OF BC13mut2

E. coli whole cells transformed with the pOTS208NS1BLmut2 plasmid [SEQ ID NO: 54] as described in Example 16 above were recovered after fermentation by centrifugation or tangential flow filtration, washed to remove media, and stored at -70°C until use.

A. Step 1: Lysis and centrifugation (Isolation)

E. coli cells, 500 gm wet cell weight (WCW), were thawed and suspended in 4-7 volumes (2L) of buffer containing 0.025 M Tris-HCl, 0.005 M EDTA, pH 8.0. Chicken egg lysozyme (Calbiochem; suspension at 100 mg/mL) was added to a final concentration of 1 g/L and the preparation stirred with a Tekmar mixer at room temperature for 1 hour.

The lysate was centrifuged at 15,000 x g for 1 hour at 4°C and the supernatant discarded. The pellet (P1) was resuspended in 5 mL per gram of original wet cell weight of buffer consisting of 0.025 M Tris-HCl, 0.002 M $MgCl_2$, pH 8.0 (about 2.5L).

The yield of this step was 90-100% by SDS-PAGE analysis, and 65-100% by RP-HPLC for product.

B. Step 2: Nuclease digestion and extraction

The preparation was treated with benzonase to digest nucleic acids, then extracted with nonionic detergents to reduce the levels of *E. coli* contaminants in the pellet. Benzon nuclease, 0.2 mL per L of suspension, was added to the suspension, which was then stirred at room temperature for 1 hr. The sample was diluted with one volume of cold water containing 2% w/v Triton X-100 and 0.2% deoxycholate and stirred for 30 min at or below 15°C. Centrifugation was repeated as in step 1 and the supernatant discarded.

C. Step 3: Urea extinction

The pellet (P2) was extracted with 5 mL/gm WCW of cold 0.025 M NaH_2PO_4, 0.025 M Tris-HCl, pH 6.0, containing 4 M urea and 10 mM dithiothreitol (DTT). The Tekmar was used at a very low speed to mix, and temperature held below 15°C. The sample was stirred at 4°C for 1 hr. then centrifuged as in step 1. The supernatant (S3) was discarded. The pellet (P3) must be stored in the freezer until use.

D. Step 4: Solubilization, reduction, and DEAE chromatography

The P3 pellet was solubilized and applied to anion exchange chromatography. This step removes remaining nucleic acid and major host cell proteins. P3 was suspended to 5 mL per gm WCW in .01 M Tris base, 8M urea (pH

not adjusted). DTT was added to 25 mM. The pH was then adjusted to 12.5 using 6N NaOH, stirring for 15 min at room temperature, immediately followed by a 5-fold dilution of the same with 10 mM boric acid containing 25 mM DTT. If needed, the sample may be diluted to keep conductivity below 2mS/cm. The pH was adjusted to 9.0 and the sample stirred for up to 2 hour at room temperature.

The pH 12.5 treatment was necessary to complete solubilization of the B/Lee protein. However since carbamylation may occur under these conditions, the time was controlled very carefully. In addition, the pH 9 adjusted sample was unstable and cannot be held.

The sample (no more than 12 mg total protein per mL of resin) was then loaded onto a 14 x 250 cm (0.75L) DEAE Toyopearl 650M column equilibrated with buffer A. All steps were performed at room temperature at a linear velocity of 100 cm/hr. The column was washed sequentially with 2-3 column volumes of buffers B, C, and D, then eluted with buffer E. When protein began to elute from the column, flow was stopped for 15-20 minutes to improve the efficiency of elution of the B/Lee product; then the peak of product protein was collected. The column was cleaned with buffer D followed by 0.5 N NaOH.

The yield of this step was 85-90% by SDS-PAGE or Western blot analysis, and was estimated at 65-70% by RP-HPLC assay for product.

E. <u>Step 5: Pretreatment and reverse phase chromatography</u>

The buffer E eluate from step 4 was adjusted to no more than 1 g/L protein concentration and made 2% in SDS, 30 mM DTT, 0.1% M Tris, 5 mM EDTA, pH 9, then heated at either: 90°C for 60, 95°C for 30 min, or 100°C for 25 minutes, using a heat exchanger or water bath. This treatment was necessary to break up aggregates and prepare the sample for RP chromatography. The sample was cooled to room temperature and 2-propanol was added to 10% v/v.

The sample was injected on an Amberchrome reverse phase column equilibrated in 10% 2-propanol/0.2% trifluoroacetic acid (TFA)/water. The gradient shown in Table 1 was used to elute the column. Fractions containing product were analyzed by analytical RP-HPLC, pooled, and held at 4°C. The column was 25cm in height and was run at a linear velocity of 75-80 cm/hr at ambient temperature. An Amicon Vantage column, 9 cm in diameter, was used. The loading capacity of the column was 2 g/L.

The reverse phase column step has a yield of 30-80% (60-80% is typical).

F. Step 6: Precipitation

The pH of the RP eluate was adjusted to 6.0 +/- 0.5 using 1 N NaOH. After 10-15 min of stirring at room temperature, the precipitate was collected by centrifugation at 16,000 x g for 30 min at 4°C. The precipitate was resuspended to approximately 6-8 mg/mL protein concentration in 25 mM Tris, 8 M urea. DTT was added to 25 mM, and the sample stirred for 30 min at room temperature. The pH was adjusted to 12.5 and stirring repeated for 15 min, immediately followed by pH adjustment to 9.0 using HCl.

Alternately, the precipitate was suspended in buffer containing 0.1 M Tris-HCl, 2% SDS, 0.01 M EDTA, pH 8.0-9.0. DTT was added to 25 mM, and stirred 15-30 min until the solution was clear and all of the precipitate had dissolved. The sample was immediately taken to the next step.

G. Step 7: Desalting and preparation of final product.

A 7 x 10 cm column was packed with Sephadex G25M (Pharmacia) at room temperature. It was equilibrated with 3-7 column volumes of 25 mM Tris-HCl, pH 9.,0, containing 5% w/v mannitol. Sample, at 6-10 mg/mL protein concentration, is injected on the column (20-25% of total column volume, i.e. 80-100 mL per injection). The column was developed at 150 cm/hr linear velocity and the product desalted into the column buffer. The final product can be stored at 4°C.

The yield of steps 6 and 7 together was no less than 90%.

The product of the purification process was recovered at an overall yield of about 20-40%, and was over 95% pure by SDS-PAGE and RP-HPLC analysis. The final yield is about 3 g/500 g well cell weight.

Table 4

Gradient for RP-LC of B/Lee

Time	Flow	%A	%B
0	80	90	10
5	80	90	10
20	80	55	45
120	80	35	65
145	80	10	90
180	80	10	90
181	0	10	90

A: 0.2% TFA in water

B: 99.8% 2-propanol/0.2% TFA

H. Purification of FluD, $NS1_{(1-81)}HA2_{(65-222)}$

FluD (Example 10) may be purified in much the same manner as the B/Lee with the following parameter alterations. For DEAE chromatography, the FluD column was equilibrated in 8M urea, 50 mM Tris, 25 mM borate at pH 9.0. After the sample is loaded, sequential washes are performed with the following buffers: 4M urea in Tris-borate pH 9.0, 4 M urea and 0.4 M NaCl in Tris-borate pH 9.0, and Tris-borate pH 9.0. The product is eluted with a step elution of 2% SDS, 0.1 to 0.25 M NaCl, in Tris-borate pH 9.0. Prior to RPLC, the protein concentration is adjusted to 1 mg/mL or less, the product is heated at 95°C for 30 minutes, and cooled, and 2-propanol is added to 10% v/v. The column is then loaded. RPLC is then performed on Amberchrome resin, as described above for B/Lee. Up to 2-3 mg of protein are loaded per ml of resin. The final yield is about 4 g/500 g wet cell weight.

EXAMPLE 19 - 3-PART INFLUENZA VACCINE

A recombinant vaccine was formulated to contain 1 µg each of the recombinant proteins $NS1_{(1-81)}HA2_{(65-222)}$ (Example 11), $NS1_{(1-81)}H3HA2_{(1-221)}mut5255$ (Example 10), and the BC13mut2 (described in Example 15 above) in Al^{+3} (100 µg) plus 3-o-deacylated monophosphoryl-lipid A (3D-MPL) (5 µg) [described in U.S. Patent No. 4,912,093; commercially available from Ribi Immunochem Research, Inc., Hamilton, Montana]. Prior to inclusion in the recombinant vaccine, the influenza proteins were purified as described in Example 15 above to remove any contaminating bacterial proteins, DNA, and endotoxin.

Mice (female, $CB6F_1$) were divided randomly into groups with 15 mice per group. The mice were injected subcutaneously on days 0 and 21 with the recombinant vaccine. A group of control mice were injected with the same dose of Al/MPL without antigen according to the same schedule. Mice were challenged with 3-5 LD_{50} doses of virus on day 49 and survival was monitored through day 21 post-challenge. In the following table showing these results, N.D. = not done and under the antigens, H1 = $NS1_{(1-81)}HA2_{(65-222)}$, H3 = $NS1_{(1-81)}H3HA2_{(1-221)}mut5855$ and B = $NS1_{(1-81)}BLHA2_{(1-223)}mut2$.

Table 5

Type A and B Cross-Protection in Mice Immunized with a Combination of Recombinant HA2 Antigens

Antigen		Percent Survival After Challenge with:		
		A/Pr/8/34 (H1)	A/HK/68 (H3)	B/Lee/40 (B)
#1	H1/H3/B	73*	80*	73
	H1	60*	N.D.	N.D.
	H3	N.D.	73*	0
	B	0	7	33[1]
	control	7	0	0
#2	H1/H3/B	93*	80*	100*
	H1	86*	N.D.	N.D.
	H3	N.D.	53**	N.D.
	B	N.D.	N.D.	80*
	control	0	7	7

* $p \leq 0.001$ vs. control group

** $p \leq 0.01$ vs. control group

[1] $p > 0.05$ (not statistically different than control group)

The data in Table 5 above results from two experiments that demonstrate that mice vaccinated with the combination of H1, H3, and Type B HA2 antigens were protected against all three virus challenges (H1, H3 and Type B) (>73-100% survival vs. 0-7% in controls). The H1 and H3 antigens in Al/MPL were subtype protective when administered individually as shown in Table 5. The Type B construct administered without the other antigens was only protective in one study (Exp. 1; 33% survival vs. 0% survival in controls but protected 80% of the mice in a second study, Exp. 2). Thus, preliminary data shows equivocal data on the stability of the Type B construct when formulated in Al/MPL in the absence of the other HA2 antigens. Studies are ongoing to confirm the stability of the construct in other formulations and in NIH/Swiss mice to confirm activity in an outbred system.

Although each antigen contains the $NS1_{(1-81)}$ regions from A/PR/8/34 (H1) virus, protections against H1 challenge was only achieved with the D protein which contains the H1HA2 region as well. Thus, the H3HA2 and Type B HA2 portions of each chimeric antigen are responsible for conferring subtype-

specific protection. Thus, the combined HA2 constructs provide cross-protections for all currently circulating influenza Type A (H1 and H3 subtypes) and Type B viruses.

Survival of NIH/Swiss outbred mice immunized with the mutant $NS_{(1-81)}BHA2_{(1-223)}$(met-leu) (not shown) showed activity at 100 micrograms (73% survival), but reduced activity at lower doses. This confirms earlier studies in outbred mice showing reduced potency relative to H1 or H3 constructs (which are active at $\geq$ 1 microgram per dose). In contrast, in $CB6F_1$ inbred mice, an inverse dose response or no dose response is seen with $NS_{(1-81)}BHA2_{(1-223)}$(met-leu).

EXAMPLE 20 - PLASMID pMS3H3HA

Plasmid pFV88 contains the entire 221 amino acid length HA2 from A/Udorn, an H3 subtype virus [C. J. Lai *et al.*, Proc. Natl. Acad. Sci. USA, 77:210-214 (1980)], which HA2 nucleic acid sequence is illustrated in Fig. 7 [SEQ ID NO: 1]. This plasmid was cut with Pst I. The resulting 1900 bp fragment, which contains the entire HA (HA1 and HA2) fragment and some GC tailing, was then inserted into pUC18 [Bethesda Research Laboratories]. The resulting plasmid is termed pMS3 or pMS3H3HA.

EXAMPLE 21 - pMG1

Plasmid pAPR801 is a pBR322-derived cloning vector which carries the NS1 coding region (A/PR/8/34). It is described by Young *et al.*, in The Origin of Pandemic Influenza Viruses, ed. by W. G. Laver, Elsevier Science Publishing Co. (1983).

Plasmid pAS1 is a pBR322-derived expression vector which contains the P_L promoter, an N utilization site (to relieve transcriptional polarity effects in the presence of N protein) and the cII ribosome binding site including the cII translation initiation codon followed immediately by a BamHI site. It is described by Rosenberg *et al.*, in Methods Enzymol., 101:123-138 (1983).

Plasmid pAS1ΔEH was prepared by deleting a non-essential EcoRI-HindIII region of pBR322 origin from pAS1. A 1236 base pair BamHI fragment of pAPR801, containing the NS1 coding region in 861 base pairs of viral origin and 375 base pairs of pBR322 origin, was inserted into the BamHI site of pAS1ΔEH. The resulting plasmid, pAS1ΔEH/801 expresses authentic NS1 (230 amino acids). The plasmid has an NcoI site between the codons for amino acids 81 and 82 and an NruI site 3' to the NS sequences. The BamHI site between amino acids 1 and 2 is retained.

Plasmid pMG27N, a pAS1 derivative [Mol. Cell. Biol., 5:1015-1024 (1985)], was cut with BamHI and SacI and ligated to a BamHI/NcoI fragment encoding the first 81 amino acids of NS1 from pAS1ΔEH801 and a synthetic DNA NcoI/SacI fragment of the following sequence:

SEQ ID NO: 10:

5'-CATGGATCATATGTTAACAGATATCAAGGCCTGACTGACTGAGAGCT-3'

SEQ ID NO: 58:

3'- CTAGTATACAATTGTCTATAGTTCCGGACTGACTGACTC -5'

The resulting plasmid, pMG1, allows the insertion of DNA fragments after the first 81 amino acids of NS1 in any of the three reading frames within the synthetic linker fragment followed by termination codons in all three reading frames.

EXAMPLE 22 - pMG1H3HA

Plasmid pMG1, described above in Example 21, was digested with NcoI and XbaI, releasing a 54 bp fragment, which was discarded. pMS3H3HA, described in Example 1 above, was digested with HhaI and XbaI, and a 701 bp fragment containing the coding sequence for the HA2 subunit of influenza strain A/Udorn (H3N2) was isolated, as illustrated in Fig. 1 [SEQ ID NO: 1].

Synthetic oligonucleotides were annealed to generate an NcoI 5' overhang sequence (at the 5' end) and a HhaI 3' overhang sequence (at the 3' end). The sequence of these oligonucleotides is as follows:

SEQ ID NO: 66: 5'-CATGGGCGCCCATATGGGCATATTCGGCG-3'

SEQ ID NO: 67: 3'- CCGCGGGTATACCCGTATAAGCC -5'

The annealing reaction was performed as follows. The annealing mixture was made up of 2.5μL each of 5' oligo (1.3 μg/μL), the 3' oligo (1.2 μg/μL), and added water (15 μL) to a final volume of 20 μL. The reaction tubes were then placed in 4 mL culture tubes containing water which had been heated to 65°C for 10 minutes and allowed to cool down slowly. The tubes were then put on ice and used immediately for ligation.

This three part ligation generates pMG1H3HA2$_{(1-221)}$ [SEQ ID NO: 9] which codes for the first 81 amino acids of NS1 fused to four amino acids donated from the linker and amino acids 1-221 of the HA2 subunit. This sequence is illustrated in Fig. 2 [SEQ ID NO: 9 & 10]. This molecule is also designated NS1$_{(1-81)}$H3HA2$_{(1-221)}$ [SEQ ID NO: 9 & 10].

EXAMPLE 23 - PREPARING SEED VIRUS AND RAISING ANTISERA

The seed virus, A/Udorn, was prepared according to the procedures described in P. Palese and J. Schulman, Virol., 57:227-237 (1974). Briefly, this technique is as follows.

Influenza virus strain A/Udorn was inoculated in 10-day old embryonated hen's eggs into the allantoic cavity. The eggs were incubated for 24-48 hours at 35°C then chilled at 4°C overnight. A portion of the eggshell over the airsac was removed and the allantoic fluid was aseptically removed using a 10-ml syringe. The fluid was centrifuged at low speed (3,000 x g) to remove particulates. This clarified supernatant was centrifuged at high speed using an SW28 Beckman rotor at 27,000 rpm (4°C for 90 minutes), resulting in the virus pellet. The virus was resuspended in 10 mM Tris (pH 7.5) containing 100 mM NaCl, 1 mM EDTA and repelleted as before. The virus was layered on 30-60% sucrose gradient in 1 mM EDTA (NTE) and spun for 3-5 hours at 25,000 rpm. The band in the middle of the tube was withdrawn, diluted in NTE and centrifuged at 27,000 rpm for 90 minutes. The pellet was suspended in phosphate-buffered saline (PBS). These viral particles were used as immunogens for preparation of antisera.

Antisera was prepared as follows. 100-200 micrograms of purified virus in complete Freund's adjuvant was injected into the subscapula of a New Zealand White rabbit. A second injection in incomplete Freund's adjuvant was done 4 weeks later, and the animals were bled 7-10 days later.

EXAMPLE 24 - MODIFICATION AND EXPRESSION OF H3HA2 FUSION PROTEINS

The modified nucleotide sequences encoding the H3HA2 proteins were prepared by mutating the nucleotide sequences of the fusion proteins prepared according to Example 22 above. Site directed mutagenesis using the Altered Sites System [Promega Corporation] according to the manufacturer's directions was used to change nucleotide numbers, 622, 625 and 634 (A to C) and 624, 627, and 636 (G to T) of nucleotide sequences [SEQ ID NO:9] encoding the $NS1_{(1-81)}H3HA2_{(1-221)}$ fusion protein of Fig. 3 [SEQ ID NO:10], thereby changing the codons at these regions from AGG to CGT, both encoding Arg. These changes correspond to nucleotide numbers 367, 370 and 379 (A to C) and 369, 372 and 381 (G to T) of the HA2 fragment of Fig. 7 [SEQ ID NO:1].

Fig. 2 illustrates the modified nucleotide sequences of the fusion proteins [SEQ ID NO: 58] by contrast with the nucleotide sequence [SEQ ID NO:9] of the "unmodified" fusion proteins (nucleotide changes below and amino acid

changes in above sequences of unmodified fusion protein). Mutagenesis on this sequence was carried out according to the method provided with the pSelect kit from Promega.

A. $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO: 10]

Briefly, cloning for the mutagenesis was performed as follows. The pSelect plasmid [Promega] and pMG1H3HA2 (Example 22) were each digested with HindIII. These two plasmids were ligated together and selected on tetracycline plates. The resulting vector is pSelH3HA2. Mutagenesis was performed according to Promega's kit. The following oligonucleotide was used: SEQ ID NO:68: 5'-AAACTGTTTG AAAAAACACG TCGTCAACTG CGTGAAAATG CTGACGACAT GGGC -3'.

Clones were verified by restriction endonuclease HincII. The resulting plasmid, pSelH3HA2mut5585 was digested with NcoI and XbaI, and a 748 bp fragment coding for the H3HA2mut5585 polypeptide was isolated. pOTS208NS181 (Eco-740) was digested with NcoI and XbaI. The ligation of linear pOTS208NS181Nco and the 748 bp fragment resulted in pOTS208NS1H3mut5585 [SEQ ID NO:58]. This vector codes for the polypeptide, $NS1_{(1-81)}H3HA2_{(1-221)}$ [SEQ ID NO:10].

B. Expression of mutated $NS1_{(1-81)}H3HA2$ proteins The plasmid of A was transfected into *E. coli* strain AR58 [SmithKline Beecham]. Cultures are grown at 32°C to mid-log phase at which time cultures are shifted to 39.5°C for two hours. The *E. coli* cell pellets containing the recombinant polypeptide are then stored at -70°C until used. Production of the $NS1_{(1-81)}H3HA2_{(1-221)}$ protein [SEQ ID NO:10] is confirmed by Western blot analysis [Towbin *et al.*, Proc. Natl. Acad. Sci. U.S.A., 76:4350 (1979)] using antisera prepared against A/Udorn virus, as described in Example 23. A major immunoreactive species is expected at a molecular weight of approximately 35,00 daltons.

The expression levels obtained are about 50-100% higher than those obtained by the expression of the unmodified coding sequences in the same expression system.

EXAMPLE 25 - tRNA INSERTION INTO HOST CELLS EXPRESSING H3 PROTEIN

E. coli host cells containing H3N2 fusion protein obtained as described in Example 22 above were transformed using conventional techniques. See, e.g. Sambrook et al, cited above.

Briefly, a culture of *E. coli* strain MM294cI$^+$ containing the plasmid pDC952 was grown overnight in Luria broth with chloramphenicol. The plasmid pDC952 carries the *argU* gene which encodes the tRNA that recognizes the AGA/AGG codons [P. Saxena and J. Walker, J. Bacteriol., 174(6):1956-1964 (Mar. 1992)]. From this culture the plasmid pDC952 was prepared. A second culture of *E. coli*, strain AR13 [SmithKline Beecham] carrying the plasmid for the H3 flu antigen, was grown overnight in Luria broth with kanamycin. These cells were made competent for transformation as described below.

The H3/AR13 overnight culture was diluted 1:50 in LB and kanamycin (50 mL total) and incubated at 37°C until it reached an $O.D._{650}$ of 0.6. The culture was then transferred to a 50 mL conical tube and chilled at about 4°C. Following this, the tube was centrifuged in a TJ6 centrifuge (10 min; 2000-3000 rev/min), the pellet resuspended in 25 mL 100 mM $CaCl_2$, and placed on ice for about 30 minutes. The pellet was then centrifuged as described above and resuspended in about 2.5 mL 100 mM $CaCl_2$.

The competent cells were aliquoted (100 µl) into three separate sterile tubes. The first tube was the negative control and did not receive any DNA. The second tube was a positive control and 1 µl of plasmid pT7II was added to the cells. To the third tube was added 3 µl of pDC952. These controls served to ensure that transformation occurred. Each tube of cells was mixed, placed on ice for 60 min., heat shocked at 37°C in a water bath for 2 minutes, and incubated in a 32°C water bath for 60 min. after adding 1 mL LB. The tubes were then microfuged for 1 minute and the supernatants poured off until only about 200 µL were left. The pellets were then resuspended in the remaining supernatant and plated as follows: (1) on LB and chloramphenicol, (2) on LB and ampicillin, and (3) on LB and chloramphenicol and kanamycin. The plates were then incubated at 32°C overnight.

Shake flasks were inoculated with the control strain, H3/AR13, and 4 transformants, pDC952/H3/AR13, and grown at 32°C to an optical density of 0.6 to 0.7 at which point the cultures were shifted to 39.5°C for 3 hours. Samples were taken at induction start (temperature shift to 39.5°C) and 3 hours post-induction. These samples were analyzed by high performance liquid chromatography (HPLC) and Western blotting.

The results of these analyses indicated that expression of H3 had increased by as much as 80% and the presence of the *argU* gene had eliminated the lowest western positive band as compared with the wild-type constructs (H3/AR13). It is believed that these results were obtained by eliminating the frameshifting caused by tandem AGG rare arginine codons. Further, there did not appear to be any difference in product quality between the H3 mutant prepared according to Example 24, and the argU tRNA transformants made according to this Example.

Numerous modifications and variations of the present invention are included in the above-identified specification and are expected to be obvious to one of skill in the art. Such modifications and alterations to the compositions and processes of the present invention are believed to be encompassed in the scope of the claims appended hereto.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Shatzman, Allan
Scott, Miller
Dillon, Susan B.
Kane, James

(ii) TITLE OF INVENTION: Vaccinal Polypeptides

(iii) NUMBER OF SEQUENCES: 72

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SmithKline Beecham Corporation - Corporate Patents
(B) STREET: U.S. Mailcode UW2220 - 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: Pennsylvania
(E) COUNTRY: USA
(F) ZIP: 19406-2799

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 149,150
(B) FILING DATE: 05-NOV-1993

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 013,415
(B) FILING DATE: 01-FEB-1993

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 108,914
(B) FILING DATE: 18-AUG-1993

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 837,773
(B) FILING DATE: 18-FEB-1992

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 751,896
(B) FILING DATE: 30-AUG-1991

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 387,200
(B) FILING DATE: 28-JUL-1989

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 238,801
(B) FILING DATE: 02-NOV-1988

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 645,732
(B) FILING DATE: 30-AUG-1984

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Baumeister, Kirk
(B) REGISTRATION NUMBER: 33,833
(C) REFERENCE/DOCKET NUMBER: P50134 PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 215-270-5096
(B) TELEFAX: 215-270-5090

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 666 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..663

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GGC ATA TTC GGC GCA ATA GCA GGT TTC ATA GAA AAT GGT TGG GAG GGA      48
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
  1               5                  10                  15

ATG ATA GAC GGT TGG TAC GGT TTC AGG CAT CAA AAT TCT GAG GGC ACA      96
Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
             20                  25                  30

GGA CAA GCA GCA GAT CTT AAA AGC ACT CAA GCA GCC ATC GAC CAA ATC     144
Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
         35                  40                  45

AAT GGG AAA CTG AAT AGG GTA ATC GAG AAG ACG AAC GAG AAA TTC CAT     192
Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
     50                  55                  60
```

CAA ATC GAA AAG GAA TTC TCA GAA GTA GAA GGG AGA ATT CAG GAC CTC 240
Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
65 70 75 80

GAG AAA TAC GTT GAA GAC ACT AAA ATA GAT CTC TGG TCT TAC AAT GCG 288
Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
85 90 95

GAG CTT CTT GTC GCT CTG GAG AAC CAA CAT ACA ATT GAT CTG ACT GAC 336
Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
100 105 110

TCG GAA ATG AAC AAA CTG TTT GAA AAA ACA AGG AGG CAA CTG AGG GAA 384
Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
115 120 125

AAT GCT GAG GAC ATG GGC AAT GGT TGC TTC AAA ATA TAC CAC AAA TGT 432
Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
130 135 140

GAC AAT GCT TGC ATA GGG TCA ATC AGA AAT GGG ACT TAT GAC CAT GAT 480
Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145 150 155 160

GTA TAC AGA GAC GAA GCA TTA AAC AAC CGG TTT CAG ATC AAA GGT GTT 528
Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
165 170 175

GAA CTG AAG TCA GGA TAC AAA GAC TGG ATC CTG TGG ATT TCC TTT GCC 576
Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala
180 185 190

ATA TCA TGC TTT TTG CTT TGT GTT GTT TTG CTG GGG TTC ATC ATG TGG 624
Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
195 200 205

GCC TGC CAG AAA GGC AAC ATT AGG TGC AAC ATT TGC ATT TGA 666
Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
210 215 220

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
1 5 10 15

```
Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
            20                  25                  30

Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
        35                  40                  45

Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
    50                  55                  60

Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
65                  70                  75                  80

Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                85                  90                  95

Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala
            180                 185                 190

Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
        195                 200                 205

Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 666 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
- (A) NAME/KEY: CDS
- (B) LOCATION: 1..663

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
GGC ATA TTC GGC GCA ATA GCA GGT TTC ATA GAA AAT GGT TGG GAG GGA      48
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
 1               5                  10                  15

ATG ATA GAC GGT TGG TAC GGT TTC AGG CAT CAA AAT TCC GAG GGC ACA      96
Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
            20                  25                  30

GGA CAA GCA GCA GAT CTT AAA AGC ACT CAA GCA GCC ATC GAC CAA ATC     144
Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
        35                  40                  45

AAT GGG AAA CTG AAT AGG GTA ATC GAG AAG ACG AAC GAG AAA TTC CAT     192
Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
    50                  55                  60

CAA ATC GAA AAG GAA TTC TCA GAA GTA GAA GGG AGA ATT CAG GAC CTC     240
Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
65                  70                  75                  80

GAG AAA TAC GTT GAA GAC ACT AAA ATA GAT CTC TGG TCT TAC AAT GCG     288
Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                85                  90                  95

GAG CTT CTT GTC GCT CTG GAG AAC CAA CAT ACA ATT GAT CTG ACT GAC     336
Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

TCG GAA ATG AAC AAA CTG TTT GAA AAA ACA AGG AGG CAA CTG AGG GAA     384
Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

AAT GCT GAG GAC ATG GGC AAT GGT TGC TTC AAA ATA TAC CAC AAA TGT     432
Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

GAC AAT GCT TGC ATA GGG TCA ATC AGA AAT GGG ACT TAT GAC CAT GAT     480
Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

GTA TAC AGA GAC GAA GCA TTA AAC AAC CGG TTT CAG ATC AAA GGT GTT     528
Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

GAA CTG AAG TCA GGA TAC AAA GAC TGG ATC CTG TGG ATT TCC TTT GCC     576
Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala
            180                 185                 190

ATA TCA TGC TTT TTG CTT TGT GTT GTT TTG CTG GGG TTC ATC ATG TGG     624
Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
        195                 200                 205
```

```
GCC TGC CAA AAA GGC AAC ATT AGG TGC AAC ATT TGC ATT TGA                666
Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
  1               5                  10                  15

Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
             20                  25                  30

Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
         35                  40                  45

Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
     50                  55                  60

Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
 65                  70                  75                  80

Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                 85                  90                  95

Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala
            180                 185                 190

Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
        195                 200                 205
```

```
Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220

(2) INFORMATION FOR SEQ ID NO:5:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 670 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: unknown

    (ii) MOLECULE TYPE: DNA (genomic)

    (ix) FEATURE:
          (A) NAME/KEY: CDS
          (B) LOCATION: 1..666

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGT CTA TTT GGA GCC ATT GCC GGT TTT ATT GAA GGG GGA TGG ACT GGA           48
Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp Thr Gly
  1               5                  10                  15

ATG ATA GAT GGA TGG TAC GGT TAT CAT CAT CAG AAT GAA CAG GGA TCA           96
Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn Glu Gln Gly Ser
             20                  25                  30

GGC TAT GCA GCG GAT CAA AAA AGC ACA CAA AAT GCC ATT AAC GGG ATT          144
Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala Ile Asn Gly Ile
         35                  40                  45

ACA AAC AAG GTG AAC TCT GTT ATC GAG AAA ATG AAC ATT CAA TTC ACA          192
Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn Ile Gln Phe Thr
     50                  55                  60

GCT GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG ATG GAA AAT TTA          240
Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu
 65                  70                  75                  80

AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG ACA TAT AAT GCA          288
Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala
                 85                  90                  95

GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG GAT TTC CAT GAC          336
Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp
            100                 105                 110

TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC CAA TTA AAG AAT          384
Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn
        115                 120                 125

AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC TAC CAC AAG TGT          432
Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys
    130                 135                 140
```

GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT TAT GAT TAT CCC 480
Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro
145 150 155 160

AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG GTA GAT GGA GTG 528
Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val
165 170 175

AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG ATC TAC TCA ACT 576
Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr
180 185 190

GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG GCA ATC AGT TTC 624
Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe
195 200 205

TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA TGC ATC 666
Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
210 215 220

TGAG 670

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 222 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp Thr Gly
1 5 10 15

Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn Glu Gln Gly Ser
20 25 30

Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala Ile Asn Gly Ile
35 40 45

Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn Ile Gln Phe Thr
50 55 60

Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu
65 70 75 80

Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala
85 90 95

```
Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp
            100                 105                 110

Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn
        115                 120                 125

Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys
    130                 135                 140

Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro
145                 150                 155                 160

Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val
                165                 170                 175

Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr
            180                 185                 190

Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe
        195                 200                 205

Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 670 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..670

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
GGCATATTCG GCGCAATAGC AGGTTTCATA GAAAATGGTT GGGAGGGAAT GATAGACGGT    60

TGGTACGGTT TCAGGCATCA AAATTCNGAG GGCACAGGAC AAGCAGCAGA TCTTAAAAGC   120

ACTCAAGCAG CCATCGACCA AATCAATGGG AAACTGAATA GGGTAATCGA GAAGACGAAC   180

GAGAAATTCC ATCAAATCGA AAAGGAATTC TCAGAAGTAG AAGGGAGAAT TCAGGACCTC   240

GAGAAATACG TTGAAGACAC TAAAATAGAT CTCTGGTCTT ACAATGCGGA GCTTCTTGTC   300

GCTCTGGAGA ACCAACATAC AATTGATCTG ACTGACTCGG AAATGAACAA ACTGTTTGAA   360

AAAACAAGGA GGCAACTGAG GGAAAATGCT GAGGACATGG GCAATGGTTG CTTCAAAATA   420

TACCACAAAT GTGACAATGC TTGCATAGGG TCAATCAGAA ATGGGACTTA TGACCATGAT   480
```

GTATACAGAG ACGAAGCATT AAACAACCGG TTTCAGATCA AAGGTGTTGA ACTGAAGTCA 540

GGATACAAAG ACTGGATCCT GTGGATTTCC TTTGCCATAT CATGCTTTTT GCTTTGTGTT 600

GTTTTGCTGG GGTTCATCAN NNTGTGGGCC TGCCANAAAG GCAACATTAG GTGCAACATT 660

TGCATTTGAN 670

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 222 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
1               5                   10                  15

Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
            20                  25                  30

Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
        35                  40                  45

Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
    50                  55                  60

Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
65                  70                  75                  80

Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                85                  90                  95

Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

Glu Leu Lys Ser Xaa Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe
            180                 185                 190
```

```
Ala Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met
        195                 200                 205

Trp Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 918 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..918

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG      48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC      96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC     144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA     192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC     240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

ATG GGC GCC CAT ATG GGC ATA TTC GGC GCA ATA GCA GGT TTC ATA GAA     288
Met Gly Ala His Met Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu
                 85                  90                  95

AAT GGT TGG GAG GGA ATG ATA GAC GGT TGG TAC GGT TTC AGG CAT CAA     336
Asn Gly Trp Glu Gly Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln
            100                 105                 110
```

AAT TCT GAG GGC ACA GGA CAA GCA GCA GAT CTT AAA AGC ACT CAA GCA 384
Asn Ser Glu Gly Thr Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala
115 120 125

GCC ATC GAC CAA ATC AAT GGG AAA CTG AAT AGG GTA ATC GAG AAG ACG 432
Ala Ile Asp Gln Ile Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr
130 135 140

AAC GAG AAA TTC CAT CAA ATC GAA AAG GAA TTC TCA GAA GTA GAA GGG 480
Asn Glu Lys Phe His Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly
145 150 155 160

AGA ATT CAG GAC CTC GAG AAA TAC GTT GAA GAC ACT AAA ATA GAT CTC 528
Arg Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu
165 170 175

TGG TCT TAC AAT GCG GAG CTT CTT GTC GCT CTG GAG AAC CAA CAT ACA 576
Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr
180 185 190

ATT GAT CTG ACT GAC TCG GAA ATG AAC AAA CTG TTT GAA AAA ACA AGG 624
Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg
195 200 205

AGG CAA CTG AGG GAA AAT GCT GAG GAC ATG GGC AAT GGT TGC TTC AAA 672
Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys
210 215 220

ATA TAC CAC AAA TGT GAC AAT GCT TGC ATA GGG TCA ATC AGA AAT GGG 720
Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly
225 230 235 240

ACT TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA AAC AAC CGG TTT 768
Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe
245 250 255

CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA TAC AAA GAC TGG ATC CTG 816
Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu
260 265 270

TGG ATT TCC TTT GCC ATA TCA TGC TTT TTG CTT TGT GTT GTT TTG CTG 864
Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu
275 280 285

GGG TTC ATC ATG TGG GCC TGC CAA AAA GGC AAC ATT AGG TGC AAC ATT 912
Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile
290 295 300

TGC ATT 918
Cys Ile
305

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 306 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gly Ala His Met Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu
                85                  90                  95

Asn Gly Trp Glu Gly Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln
            100                 105                 110

Asn Ser Glu Gly Thr Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala
        115                 120                 125

Ala Ile Asp Gln Ile Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr
    130                 135                 140

Asn Glu Lys Phe His Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly
145                 150                 155                 160

Arg Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu
                165                 170                 175

Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr
            180                 185                 190

Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg
        195                 200                 205

Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys
    210                 215                 220

Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly
225                 230                 235                 240
```

Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe
245 250 255

Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu
260 265 270

Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu
275 280 285

Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile
290 295 300

Cys Ile
305

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 690 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..690

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG 48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1 5 10 15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC 96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
20 25 30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC 144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
35 40 45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA 192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
50 55 60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC 240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65 70 75 80

```
ATG GAT CAT ATG TTA ATT CAG GAC CTC GAG AAA TAC GTT GAA GAC ACT     288
Met Asp His Met Leu Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr
                85                  90                  95

AAA ATA GAT CTC TGG TCT TAC AAT GCG GAG CTT CTT GTC GCT CTG GAG     336
Lys Ile Asp Leu Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu
            100                 105                 110

AAC CAA CAT ACA ATT GAT CTG ACT GAC TCG GAA ATG AAC AAA CTG TTT     384
Asn Gln His Thr Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe
        115                 120                 125

GAA AAA ACA AGG AGG CAA CTG AGG GAA AAT GCT GAG GAC ATG GGC AAT     432
Glu Lys Thr Arg Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn
    130                 135                 140

GGT TGC TTC AAA ATA TAC CAC AAA TGT GAC AAT GCT TGC ATA GGG TCA     480
Gly Cys Phe Lys Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser
145                 150                 155                 160

ATC AGA AAT GGG ACT TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA     528
Ile Arg Asn Gly Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu
                165                 170                 175

AAC AAC CGG TTT CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA TAC AAA     576
Asn Asn Arg Phe Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys
            180                 185                 190

GAC TGG ATC CTG TGG ATT TCC TTT GCC ATA TCA TGC TTT TTG CTT TGT     624
Asp Trp Ile Leu Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys
        195                 200                 205

GTT GTT TTG CTG GGG TTC ATC ATG TGG GCC TGC CAA AAA GGC AAC ATT     672
Val Val Leu Leu Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile
    210                 215                 220

AGG TGC AAC ATT TGC ATT                                             690
Arg Cys Asn Ile Cys Ile
225                 230
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 230 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1               5                   10                  15
```

```
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Asp His Met Leu Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr
                85                  90                  95

Lys Ile Asp Leu Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu
            100                 105                 110

Asn Gln His Thr Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe
        115                 120                 125

Glu Lys Thr Arg Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn
    130                 135                 140

Gly Cys Phe Lys Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser
145                 150                 155                 160

Ile Arg Asn Gly Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu
                165                 170                 175

Asn Asn Arg Phe Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys
            180                 185                 190

Asp Trp Ile Leu Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys
        195                 200                 205

Val Val Leu Leu Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile
    210                 215                 220

Arg Cys Asn Ile Cys Ile
225                 230
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 699 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

```
(ix) FEATURE:
      (A) NAME/KEY: CDS
      (B) LOCATION: 1..699

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TCC TTT CTT TGG     48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Ser Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC     96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC ATG CAT GGA TCA TAT GTT    144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met His Gly Ser Tyr Val
         35                  40                  45

AAC AAG ACA CAA GAA GCT ATA AAC AAG ATA ACA AAA AAT CTC AAC TAT    192
Asn Lys Thr Gln Glu Ala Ile Asn Lys Ile Thr Lys Asn Leu Asn Tyr
     50                  55                  60

TTA AGT GAG CTA GAA GTA AAA AAC CTT CAA AGA CTA AGC GGA GCA ATG    240
Leu Ser Glu Leu Glu Val Lys Asn Leu Gln Arg Leu Ser Gly Ala Met
 65                  70                  75                  80

AAT GAG CTT CAC GAC GAA ATA CTC GAG CTA GAC GAA AAA GTG GAT GAT    288
Asn Glu Leu His Asp Glu Ile Leu Glu Leu Asp Glu Lys Val Asp Asp
                 85                  90                  95

CTA AGA GCT GAT ACA ATA AGC TCA CAA ATA GAG CTT GCA GTC TTG CTT    336
Leu Arg Ala Asp Thr Ile Ser Ser Gln Ile Glu Leu Ala Val Leu Leu
            100                 105                 110

TCC AAC GAA GGG ATA ATA AAC AGT GAA GAT GAG CAT CTC TTG GCA CTT    384
Ser Asn Glu Gly Ile Ile Asn Ser Glu Asp Glu His Leu Leu Ala Leu
        115                 120                 125

GAA AGA AAA CTG AAG AAA ATG CTT GGC CCC TCT GCT GTA GAA ATA GGG    432
Glu Arg Lys Leu Lys Lys Met Leu Gly Pro Ser Ala Val Glu Ile Gly
    130                 135                 140

AAT GGG TGC TTT GAA ACC AAA CAC AAA TGC AAC CAG ACT TGC CTA GAC    480
Asn Gly Cys Phe Glu Thr Lys His Lys Cys Asn Gln Thr Cys Leu Asp
145                 150                 155                 160

AGG ATA GCT GCT GGC ACC TTT AAT GCA GGA GAT TTT TCT CTT CCC ACT    528
Arg Ile Ala Ala Gly Thr Phe Asn Ala Gly Asp Phe Ser Leu Pro Thr
                165                 170                 175

TTT GAT TCA TTA AAC ATT ACT GCT GCA TCT TTA AAT GAT GAT GGC TTG    576
Phe Asp Ser Leu Asn Ile Thr Ala Ala Ser Leu Asn Asp Asp Gly Leu
            180                 185                 190
```

```
GAT AAT CAT ACT ATA CTG CTC TAC TAC TCA ACT GCT GCT TCT AGC TTG        624
Asp Asn His Thr Ile Leu Leu Tyr Tyr Ser Thr Ala Ala Ser Ser Leu
        195                 200                 205

GCT GTA ACA TTA ATG ATA GCT ATC TTC ATT GTC TAC ATG GTC TCC AGA        672
Ala Val Thr Leu Met Ile Ala Ile Phe Ile Val Tyr Met Val Ser Arg
    210                 215                 220

GAC AAT GTT TCT TGT TCC ATC TGT CTG                                    699
Asp Asn Val Ser Cys Ser Ile Cys Leu
225                 230
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 233 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Ser Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met His Gly Ser Tyr Val
        35                  40                  45

Asn Lys Thr Gln Glu Ala Ile Asn Lys Ile Thr Lys Asn Leu Asn Tyr
    50                  55                  60

Leu Ser Glu Leu Glu Val Lys Asn Leu Gln Arg Leu Ser Gly Ala Met
65                  70                  75                  80

Asn Glu Leu His Asp Glu Ile Leu Glu Leu Asp Glu Lys Val Asp Asp
                85                  90                  95

Leu Arg Ala Asp Thr Ile Ser Ser Gln Ile Glu Leu Ala Val Leu Leu
            100                 105                 110

Ser Asn Glu Gly Ile Ile Asn Ser Glu Asp Glu His Leu Leu Ala Leu
        115                 120                 125

Glu Arg Lys Leu Lys Lys Met Leu Gly Pro Ser Ala Val Glu Ile Gly
    130                 135                 140

Asn Gly Cys Phe Glu Thr Lys His Lys Cys Asn Gln Thr Cys Leu Asp
145                 150                 155                 160
```

```
Arg Ile Ala Ala Gly Thr Phe Asn Ala Gly Asp Phe Ser Leu Pro Thr
                165                 170                 175

Phe Asp Ser Leu Asn Ile Thr Ala Ala Ser Leu Asn Asp Asp Gly Leu
            180                 185                 190

Asp Asn His Thr Ile Leu Leu Tyr Tyr Ser Thr Ala Ala Ser Ser Leu
        195                 200                 205

Ala Val Thr Leu Met Ile Ala Ile Phe Ile Val Tyr Met Val Ser Arg
    210                 215                 220

Asp Asn Val Ser Cys Ser Ile Cys Leu
225                 230
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 924 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..921

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG      48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC      96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC     144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA     192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC     240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

ATG GAT CTG TCC AGA GGT CTA TTT GGA GCC ATT GCC GGT TTT ATT GAA     288
Met Asp Leu Ser Arg Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu
                 85                  90                  95
```

```
GGG GGA TGG ACT GGA ATG ATA GAT GGA TGG TAC GGT TAT CAT CAT CAG      336
Gly Gly Trp Thr Gly Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln
            100                 105                 110

AAT GAA CAG GGA TCA GGC TAT GCA GCG GAT CAA AAA AGC ACA CAA AAT      384
Asn Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn
        115                 120                 125

GCC ATT AAC GGG ATT ACA AAC AAG GTG AAC TCT GTT ATC GAG AAA ATG      432
Ala Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met
    130                 135                 140

AAC ATT CAA TTC ACA GCT GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA      480
Asn Ile Gln Phe Thr Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys
145                 150                 155                 160

AGG ATG GAA AAT TTA AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT      528
Arg Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile
                165                 170                 175

TGG ACA TAT AAT GCA GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT      576
Trp Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr
            180                 185                 190

CTG GAT TTC CAT GAC TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA      624
Leu Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys
        195                 200                 205

AGC CAA TTA AAG AAT AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG      672
Ser Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu
    210                 215                 220

TTC TAC CAC AAG TGT GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG      720
Phe Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly
225                 230                 235                 240

ACT TAT GAT TAT CCC AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA      768
Thr Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu
                245                 250                 255

AAG GTA GAT GGA GTG AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG      816
Lys Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu
            260                 265                 270

GCG ATC TAC TCA ACT GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG      864
Ala Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu
        275                 280                 285

GGG GCA ATC AGT TTC TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA      912
Gly Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg
    290                 295                 300
```

ATA TGC ATC TGA 924
Ile Cys Ile
305

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 307 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

Met Asp Leu Ser Arg Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu
                 85                  90                  95

Gly Gly Trp Thr Gly Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln
            100                 105                 110

Asn Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn
        115                 120                 125

Ala Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met
    130                 135                 140

Asn Ile Gln Phe Thr Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys
145                 150                 155                 160

Arg Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile
                165                 170                 175

Trp Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr
            180                 185                 190
```

Leu Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys
195 200 205

Ser Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu
210 215 220

Phe Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly
225 230 235 240

Thr Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu
245 250 255

Lys Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu
260 265 270

Ala Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu
275 280 285

Gly Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg
290 295 300

Ile Cys Ile
305

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 729 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..726

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG 48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1 5 10 15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC 96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
20 25 30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC 144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
35 40 45

```
ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA      192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC      240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

ATG CAG ATC CCG GCT GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG      288
Met Gln Ile Pro Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
                 85                  90                  95

ATG GAA AAT TTA AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG      336
Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
            100                 105                 110

ACA TAT AAT GCA GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG      384
Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
        115                 120                 125

GAT TTC CAT GAC TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC      432
Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
    130                 135                 140

CAA TTA AAG AAT AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC      480
Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
145                 150                 155                 160

TAC CAC AAG TGT GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT      528
Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
                165                 170                 175

TAT GAT TAT CCC AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG      576
Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
            180                 185                 190

GTA GAT GGA GTG AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG      624
Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
        195                 200                 205

ATC TAC TCA ACT GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG      672
Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
    210                 215                 220

GCA ATC AGT TTC TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA      720
Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
225                 230                 235                 240

TGC ATC TGA                                                          729
Cys Ile
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 242 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
                85                  90                  95

Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
            100                 105                 110

Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
        115                 120                 125

Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
    130                 135                 140

Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
145                 150                 155                 160

Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
                165                 170                 175

Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
            180                 185                 190

Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
        195                 200                 205

Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
    210                 215                 220
```

```
Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
225                 230                 235                 240

Cys Ile
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 810 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..807

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG      48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC      96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC ATG GAT CTG TCC AGA GGT     144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met Asp Leu Ser Arg Gly
         35                  40                  45

CTA TTT GGA GCC ATT GCC GGT TTT ATT GAA GGG GGA TGG ACT GGA ATG     192
Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp Thr Gly Met
     50                  55                  60

ATA GAT GGA TGG TAC GGT TAT CAT CAT CAG AAT GAA CAG GGA TCA GGC     240
Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn Glu Gln Gly Ser Gly
 65                  70                  75                  80

TAT GCA GCG GAT CAA AAA AGC ACA CAA AAT GCC ATT AAC GGG ATT ACA     288
Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala Ile Asn Gly Ile Thr
                 85                  90                  95

AAC AAG GTG AAC TCT GTT ATC GAG AAA ATG AAC ATT CAA TTC ACA GCT     336
Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn Ile Gln Phe Thr Ala
            100                 105                 110

GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG ATG GAA AAT TTA AAT     384
Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu Asn
        115                 120                 125
```

```
AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG ACA TAT AAT GCA GAA        432
Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala Glu
    130                 135                 140

TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG GAT TTC CAT GAC TCA        480
Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp Ser
145                 150                 155                 160

AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC CAA TTA AAG AAT AAT        528
Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn Asn
                165                 170                 175

GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC TAC CAC AAG TGT GAC        576
Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys Asp
            180                 185                 190

AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT TAT GAT TAT CCC AAA        624
Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro Lys
        195                 200                 205

TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG GTA GAT GGA GTG AAA        672
Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val Lys
    210                 215                 220

TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG ATC TAC TCA ACT GTC        720
Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr Val
225                 230                 235                 240

GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG GCA ATC AGT TTC TGG        768
Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe Trp
                245                 250                 255

ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA TGC ATC TGA                810
Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
            260                 265
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 269 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1               5                   10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30
```

```
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met Asp Leu Ser Arg Gly
        35                  40                  45

Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp Thr Gly Met
    50                  55                  60

Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn Glu Gln Gly Ser Gly
65                  70                  75                  80

Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala Ile Asn Gly Ile Thr
                85                  90                  95

Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn Ile Gln Phe Thr Ala
            100                 105                 110

Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu Asn
        115                 120                 125

Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala Glu
    130                 135                 140

Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp Ser
145                 150                 155                 160

Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn Asn
                165                 170                 175

Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys Asp
            180                 185                 190

Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro Lys
        195                 200                 205

Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val Lys
    210                 215                 220

Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr Val
225                 230                 235                 240

Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe Trp
                245                 250                 255

Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
            260                 265
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 630 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

```
(ix) FEATURE:
     (A) NAME/KEY: CDS
     (B) LOCATION: 1..627

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG       48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC       96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC ATG GAT CAT ATG TTA ACA      144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met Asp His Met Leu Thr
        35                  40                  45

AGT ACT CGA TCT GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG ATG      192
Ser Thr Arg Ser Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met
    50                  55                  60

GAA AAT TTA AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG ACA      240
Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr
65                  70                  75                  80

TAT AAT GCA GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG GAT      288
Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp
                85                  90                  95

TTC CAT GAC TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC CAA      336
Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln
            100                 105                 110

TTA AAG AAT AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC TAC      384
Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr
        115                 120                 125

CAC AAG TGT GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT TAT      432
His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr
    130                 135                 140

GAT TAT CCC AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG GTA      480
Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val
145                 150                 155                 160

GAT GGA GTG AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG ATC      528
Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile
                165                 170                 175

TAC TCA ACT GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG GCA      576
Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala
            180                 185                 190
```

```
ATC AGT TTC TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA TGC      624
Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys
        195                 200                 205

ATC TGA                                                              630
Ile
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 209 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met Asp His Met Leu Thr
        35                  40                  45

Ser Thr Arg Ser Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met
    50                  55                  60

Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr
65                  70                  75                  80

Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp
                85                  90                  95

Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln
            100                 105                 110

Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr
        115                 120                 125

His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr
    130                 135                 140

Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val
145                 150                 155                 160

Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile
                165                 170                 175
```

```
Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala
            180                 185                 190

Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys
        195                 200                 205

Ile
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 717 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..714

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG        48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC        96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC       144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA       192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC       240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

ATG CAG ATC CCG GAA TTC AAC AAA TTA GAA AAA AGG ATG GAA AAT TTA       288
Met Gln Ile Pro Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu
                 85                  90                  95

AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG ACA TAT AAT GCA       336
Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala
            100                 105                 110
```

```
GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG GAT TTC CAT GAC      384
Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp
        115                 120                 125

TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC CAA TTA AAG AAT      432
Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn
    130                 135                 140

AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC TAC CAC AAG TGT      480
Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys
145                 150                 155                 160

GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT TAT GAT TAT CCC      528
Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro
                165                 170                 175

AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG GTA GAT GGA GTG      576
Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val
            180                 185                 190

AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG ATC TAC TCA ACT      624
Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr
        195                 200                 205

GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG GCA ATC AGT TTC      672
Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe
    210                 215                 220

TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA TGC ATC              714
Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
225                 230                 235

TGA                                                                  717
```

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 238 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45
```

```
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

Met Gln Ile Pro Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu
                 85                  90                  95

Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala
            100                 105                 110

Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp
        115                 120                 125

Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn
    130                 135                 140

Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys
145                 150                 155                 160

Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro
                165                 170                 175

Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val
            180                 185                 190

Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr
        195                 200                 205

Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe
    210                 215                 220

Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
225                 230                 235
```

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 681 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..678

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG      48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC      96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC     144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA     192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC     240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

ATG CAG ATC CCG AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG     288
Met Gln Ile Pro Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
                 85                  90                  95

ACA TAT AAT GCA GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG     336
Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
            100                 105                 110

GAT TTC CAT GAC TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC     384
Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
        115                 120                 125

CAA TTA AAG AAT AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC     432
Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
    130                 135                 140

TAC CAC AAG TGT GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT     480
Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
145                 150                 155                 160

TAT GAT TAT CCC AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG     528
Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
                165                 170                 175

GTA GAT GGA GTG AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG     576
Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
            180                 185                 190

ATC TAC TCA ACT GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG     624
Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
        195                 200                 205
```

```
GCA ATC AGT TTC TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA      672
Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
    210                 215                 220

TGC ATC TGA                                                          681
Cys Ile
225
```

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 226 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

Met Gln Ile Pro Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
                 85                  90                  95

Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
            100                 105                 110

Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
        115                 120                 125

Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
    130                 135                 140

Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
145                 150                 155                 160

Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
                165                 170                 175
```

```
Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
        180                 185                 190

Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
    195                 200                 205

Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
    210                 215                 220

Cys Ile
225
```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 158 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1               5                   10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Val Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro
                85                  90                  95

Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val
            100                 105                 110

Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr
        115                 120                 125

Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe
    130                 135                 140

Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 163 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1               5                   10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Asp Leu Ser Arg Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu
                85                  90                  95

Gly Gly Trp Thr Gly Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln
            100                 105                 110

Asn Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn
        115                 120                 125

Ala Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met
    130                 135                 140

Asn Ile Gln Phe Thr Ala Val Gly Lys Glu Phe Ser Cys Leu Thr Ala
145                 150                 155                 160

Tyr His Arg
```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 231 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1               5                   10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
                85                  90                  95

Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
            100                 105                 110

Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
        115                 120                 125

Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
    130                 135                 140

Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
145                 150                 155                 160

Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
                165                 170                 175

Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
            180                 185                 190

Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
        195                 200                 205

Ile Tyr Ser Thr Val Ala Ser Ser Gly Gly Ser Tyr Ser Met Glu His
    210                 215                 220

Phe Arg Trp Gly Lys Pro Val
225                 230
```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 225 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1               5                   10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
                85                  90                  95

Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
            100                 105                 110

Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
        115                 120                 125

Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
    130                 135                 140

Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
145                 150                 155                 160

Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
                165                 170                 175

Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
            180                 185                 190

Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
        195                 200                 205

Ile Tyr Ser Thr Val Ala Ser Ser Gly Gly Ser Tyr Ser Met Leu Val
    210                 215                 220
```

```
Asn
225
```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 912 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..912

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG       48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC       96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC      144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA      192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC      240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

ATG CAG ATC CCG GGT CTA TTT GGA GCC ATT GCC GGT TTT ATT GAA GGG      288
Met Gln Ile Pro Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly
                 85                  90                  95

GGA TGG ACT GGA ATG ATA GAT GGA TGG TAC GGT TAT CAT CAT CAG AAT      336
Gly Trp Thr Gly Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn
            100                 105                 110

GAA CAG GGA TCA GGC TAT GCA GCG GAT CAA AAA AGC ACA CAA AAT GCC      384
Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala
        115                 120                 125
```

```
ATT AAC GGG ATT ACA AAC AAG GTG AAC TCT GTT ATC GAG AAA ATG AAC      432
Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn
    130                 135                 140

ATT CAA TTC ACA GCT GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG      480
Ile Gln Phe Thr Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
145                 150                 155                 160

ATG GAA AAT TTA AAT AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG      528
Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
                165                 170                 175

ACA TAT AAT GCA GAA TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG      576
Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
            180                 185                 190

GAT TTC CAT GAC TCA AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC      624
Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
        195                 200                 205

CAA TTA AAG AAT AAT GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC      672
Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
    210                 215                 220

TAC CAC AAG TGT GAC AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT      720
Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
225                 230                 235                 240

TAT GAT TAT CCC AAA TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG      768
Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
                245                 250                 255

GTA GAT GGA GTG AAA TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG      816
Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
            260                 265                 270

ATC TAC TCA ACT GTC GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG      864
Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
        275                 280                 285

GCA ATC AGT TTC TGG ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA      912
Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
    290                 295                 300
```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 304 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gln Ile Pro Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly
                85                  90                  95

Gly Trp Thr Gly Met Ile Asp Gly Trp Tyr Gly Tyr His His Gln Asn
            100                 105                 110

Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr Gln Asn Ala
        115                 120                 125

Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn
    130                 135                 140

Ile Gln Phe Thr Ala Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg
145                 150                 155                 160

Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp
                165                 170                 175

Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu
            180                 185                 190

Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser
        195                 200                 205

Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe
    210                 215                 220

Tyr His Lys Cys Asp Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr
225                 230                 235                 240

Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys
                245                 250                 255

Val Asp Gly Val Lys Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala
            260                 265                 270
```

```
Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly
        275                 280                 285

Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile
    290                 295                 300
```

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 474 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..471

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
GTG GGT AAA GAA TTC AAC AAA TTA GAA AAA AGG ATG GAA AAT TTA AAT      48
Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu Asn
  1               5                  10                  15

AAA AAA GTT GAT GAT GGA TTT CTG GAC ATT TGG ACA TAT AAT GCA GAA      96
Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala Glu
             20                  25                  30

TTG TTA GTT CTA CTG GAA AAT GAA AGG ACT CTG GAT TTC CAT GAC TCA     144
Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp Ser
         35                  40                  45

AAT GTG AAG AAT CTG TAT GAG AAA GTA AAA AGC CAA TTA AAG AAT AAT     192
Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn Asn
     50                  55                  60

GCC AAA GAA ATC GGA AAT GGA TGT TTT GAG TTC TAC CAC AAG TGT GAC     240
Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys Asp
 65                  70                  75                  80

AAT GAA TGC ATG GAA AGT GTA AGA AAT GGG ACT TAT GAT TAT CCC AAA     288
Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro Lys
                 85                  90                  95

TAT TCA GAA GAG TCA AAG TTG AAC AGG GAA AAG GTA GAT GGA GTG AAA     336
Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val Lys
            100                 105                 110
```

```
TTG GAA TCA ATG GGG ATC TAT CAG ATT CTG GCG ATC TAC TCA ACT GTC      384
Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr Val
        115                 120                 125

GCC AGT TCA CTG GTG CTT TTG GTC TCC CTG GGG GCA ATC AGT TTC TGG      432
Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe Trp
    130                 135                 140

ATG TGT TCT AAT GGA TCT TTG CAG TGC AGA ATA TGC ATC TGA              474
Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 157 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
Val Gly Lys Glu Phe Asn Lys Leu Glu Lys Arg Met Glu Asn Leu Asn
 1               5                  10                  15

Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala Glu
            20                  25                  30

Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Phe His Asp Ser
        35                  40                  45

Asn Val Lys Asn Leu Tyr Glu Lys Val Lys Ser Gln Leu Lys Asn Asn
    50                  55                  60

Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys Asp
65                  70                  75                  80

Asn Glu Cys Met Glu Ser Val Arg Asn Gly Thr Tyr Asp Tyr Pro Lys
                85                  90                  95

Tyr Ser Glu Glu Ser Lys Leu Asn Arg Glu Lys Val Asp Gly Val Lys
            100                 105                 110

Leu Glu Ser Met Gly Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr Val
        115                 120                 125

Ala Ser Ser Leu Val Leu Leu Val Ser Leu Gly Ala Ile Ser Phe Trp
    130                 135                 140

Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CATGGATCAT ATGTTAACAG ATATCAAGGC CTGACTGACT GAGAGCT 47

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CTCAGTCAGT CAGGCCTTGA TATCTGTTAA CATATGATC 39

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CATGGGCGCC CATATGGGCA TATTCGGCG 29

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCGAATATGC CCATATGGGC GCC 23

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CATGGATCAT ATGTTAACAA GTACTCGATA TCAATGAGTG ACTGAAGCT 49

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TCAGTCACTC ATTGATATCG AGTACTTGTT AACATATGAT C 41

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AATTCGTACC TA 12

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GATCTAGGTA CG 12

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

AAACTGTTTG AAAAAACACG TCGTCAACTG CGTGAAAATG CTGACGACAT GGGC 54

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TGTGACAATG CTTGCATCGG TTCAATCCGT AATGGTACTT ATGACCATGA TG 52

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GATCCCGGGT GACTGACTGA 20

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GATCTCAGTC AGTCACCCGG 20

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TAAGGAGGAT ATAACATATG 20

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GATCCATATG TTATATCCTC CTTAAGGT 28

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GCATCGCCAT GAGTCACGAC G 21

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GGAGGATGGG AAGGACTCAT TGCAGGTTGG 30

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CTCTGCTGTA GAAATCGGTA ACGGTTGCTT TGAAACCAAA C 41

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GGTTTCTTGG AAGGTGGTTG GGAAGGTCTC ATTGCAGGTT GGCACGG 47

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GCTTTCCAAC GAAGGTATCA ATCAACAGTG AAGACGAGCA TCTCTTGG 48

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7616 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1879..2790

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

```
AATTCTCATG TTTGACAGCT TATCATCGAT AAGCTTCAGT TGAAGATATT AAGAACAGCC     60
TCGCAGATGA CGAATCATTG GGATTCCCAT CTTTTTTGTT TGTTGAAGGC GACACCATTG    120
GTTTTGCCAG AACTGTTTTC GGGCCGACCA CATCCGATCT GACAGATTTT TTAATCGGGA    180
AAGGAATGTC ATTAAGCAGT GGAGAGCGCG TTCAGATAGA GCCACTGATG AGGGGAACCA    240
CCAAAGACGA TGTTATGCAT ATGCATTTCA TCGGCCGAAC AACGGTGAAG GTAGAAGCCA    300
AGCTACCTGT ATTTGGCGAT ATATTAAAGG TCTTAGGGGC AACAGATATT GAAGGGGAGC    360
TTTTTGACTC ATTGGATATA GTCATTAAGC CAAAATTTAA AAGGGATATA AAAAAGGTTG    420
CCAAGGATAT TATTTTTAAC CCGTCACCTC AATTTTCAGA CATTAGCCTG CGGGCAAAAG    480
ATGAGGCCGG AGATATTTTA ACAGAACATT ATCTATCAGA AAAAGGCCAT CTCTCAGCGC    540
CTCTGAACAA GGTCACCAAT GCTGAGATAG CTGAAGAGAT GGCATATTGC TACGCAAGAA    600
TGAAAAGTGA TATACTGGAA TGTTTTAAAA GGCAGGTGGG CAAAGTTAAG GATTAATTAT    660
CAGGAGTAAT TATGCGGAAC AGAATCATGC CTGGTGTTTA CATAGTAATA ATTCCTTACG    720
TTATCGTAAG CATTTGCTAT CTCCTTTTCC GCCACTACAT TCCTGGTGTT TCTTTTTCAG    780
CTCATAGAGA TGGTCTTGGG GCGACATTGT CATCATATGC AGGAACCATG ATTGCAATCC    840
TGATTGCTGC CTTGACGTTT CTAATCGGAA GCAGAACGCG CCGACTGGCC AAGATTAGAG    900
AGTATGGGTA TATGACATCG GTAGTTATTG TCTATGCCCT TAGTTTTGTT GAGCTTGGAG    960
CTTTGTTTTT CTGCGGGTTA TTGCTTCTTT CCAGCATAAG CGGCTACATG ATACCCACTA   1020
TCGCCATCGG CATTGCCTCT GCATCGTTCA TTCATATATG CATCCTTGTT TTCCAACTAT   1080
ATAATTTGAC CAGAGAACAA GAATAACCCG GCCTCAGCGC CGGGTTTTCT TTGCCTCACG   1140
ATCGCCCCCA AAACACATAA CCAATTGTAT TTATTGAAAA ATAAATAGAT ACAACTCACT   1200
AAACATAGCA ATTCAGATCT CTCACCTACC AAACAATGCC CCCCTGCAAA AAATAAATTC   1260
ATATAAAAAA CATACAGATA ACCATCTGCG GTGATAAATT ATCTCTGGCG GTGTTGACAT   1320
AAATACCACT GGCGGTGATA CTGAGCACAT CAGCAGGACG CACTGACCAC CATGAAGGTG   1380
ACGCTCTTAA AAATTAAGCC CTGAAGAAGG GCAGCATTCA AAGCAGAAGG CTTTGGGGTG   1440
```

```
TGTGATACGA AACGAAGCAT TGGCCGTAAG TGCGATTCCG GATTAGCTGC CAATGTGCCA    1500

ATCGCGGGGG GTTTTCGTTC AGGACTACAA CTGCCACACA CCACCAAAGC TAACTGACAG    1560

GAGAATCCAG ATGGATGCAC AAACACGCCG CCGCGAACGT CGCGCAGAGA AACAGGCTCA    1620

ATGGAAAGCA GCAAATCCCC TGTTGGTTGG GGTAAGCGCA AAACCAGTTC CGAAAGATTT    1680

TTTTAACTAT AAACGCTGAT GGAAGCGTTT ATGCGGAAGA GGTAAAGCCC TTCCCGAGTA    1740

ACAAAAAAAC AACAGCATAA ATAACCCCGC TCTTACACAT TCCAGCCCTG AAAAAGGGCA    1800

TCAAATTAAA CCACACCTAT GGTGTATGCA TTTATTTGCA TACATTCAAT CAATTGTTAT    1860

CTAAGGAAAT ACTTACAT ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA    1911
                    Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val
                     1               5                  10

GAT TGC TTT CTT TGG CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA    1959
Asp Cys Phe Leu Trp His Val Arg Lys Arg Val Ala Asp Gln Glu Leu
            15                  20                  25

GGT GAT GCC CCA TTC CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA    2007
Gly Asp Ala Pro Phe Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu
        30                  35                  40

AGA GGA AGG GGC AGC ACC CTC GGT CTG GAC ATC GAG ACA GCC ACA CGT    2055
Arg Gly Arg Gly Ser Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg
    45                  50                  55

GCT GGA AAG CAG ATA GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG    2103
Ala Gly Lys Gln Ile Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu
60                  65                  70                  75

GCA CTT AAA ATG ACC ATG GGT TTC TTC GGA GCT ATT GCT GGT TTC TTG    2151
Ala Leu Lys Met Thr Met Gly Phe Phe Gly Ala Ile Ala Gly Phe Leu
                80                  85                  90

GAA GGT GGT TGG GAA GGT CTC ATT GCA GGT TGG CAC GGA TAC ACA TCT    2199
Glu Gly Gly Trp Glu Gly Leu Ile Ala Gly Trp His Gly Tyr Thr Ser
            95                  100                 105

CAT GGA GCA CAT GGA GTG GCA GTG GCA GCA GAC CTT AAG AGT ACA CAA    2247
His Gly Ala His Gly Val Ala Val Ala Ala Asp Leu Lys Ser Thr Gln
        110                 115                 120

GAA GCT ATA AAC AAG ATA ACA AAA AAT CTC AAC TAT TTA AGT GAG CTA    2295
Glu Ala Ile Asn Lys Ile Thr Lys Asn Leu Asn Tyr Leu Ser Glu Leu
    125                 130                 135

GAA GTA AAA AAC CTT CAA AGA CTA AGC GGA GCA ATG AAT GAG CTT CAC    2343
Glu Val Lys Asn Leu Gln Arg Leu Ser Gly Ala Met Asn Glu Leu His
140                 145                 150                 155
```

```
GAC GAA ATA CTC GAG CTA GAC GAA AAA GTG GAT GAT CTA AGA GCT GAT      2391
Asp Glu Ile Leu Glu Leu Asp Glu Lys Val Asp Asp Leu Arg Ala Asp
                160                 165                 170

ACA ATA AGC TCA CAA ATA GAG CTT GCA GTC TTG CTT TCC AAC GAA GGT      2439
Thr Ile Ser Ser Gln Ile Glu Leu Ala Val Leu Leu Ser Asn Glu Gly
            175                 180                 185

ATC ATC AAC AGT GAA GAC GAG CAT CTC TTG GCA CTT GAA AGA AAA CTG      2487
Ile Ile Asn Ser Glu Asp Glu His Leu Leu Ala Leu Glu Arg Lys Leu
        190                 195                 200

AAG AAA ATG CTT GGC CCC TCT GCT GTA GAA ATC GGT AAC GGT TGC TTT      2535
Lys Lys Met Leu Gly Pro Ser Ala Val Glu Ile Gly Asn Gly Cys Phe
    205                 210                 215

GAA ACC AAA CAC AAA TGC AAC CAG ACT TGC CTA GAC AGG ATA GCT GCT      2583
Glu Thr Lys His Lys Cys Asn Gln Thr Cys Leu Asp Arg Ile Ala Ala
220                 225                 230                 235

GGC ACC TTT AAT GCA GGA GAT TTT TCT CTT CCC ACT TTT GAT TCA TTA      2631
Gly Thr Phe Asn Ala Gly Asp Phe Ser Leu Pro Thr Phe Asp Ser Leu
                240                 245                 250

AAC ATT ACT GCT GCA TCT TTA AAT GAT GAT GGC TTG GAT AAT CAT ACT      2679
Asn Ile Thr Ala Ala Ser Leu Asn Asp Asp Gly Leu Asp Asn His Thr
            255                 260                 265

ATA CTG CTC TAC TAC TCA ACT GCT GCT TCT AGC TTG GCT GTA ACA TTA      2727
Ile Leu Leu Tyr Tyr Ser Thr Ala Ala Ser Ser Leu Ala Val Thr Leu
        270                 275                 280

ATG ATA GCT ATC TTC ATT GTC TAC ATG GTC TCC AGA GAC AAT GTT TCT      2775
Met Ile Ala Ile Phe Ile Val Tyr Met Val Ser Arg Asp Asn Val Ser
    285                 290                 295

TGT TCC ATC TGT CTG TGAGGGAGAT TAAGCCCTGT GTTTTCCTTT ACTGTAGTGC      2830
Cys Ser Ile Cys Leu
300

TCATTTGCTT GTCACCATTA CAAAGAAACG TTATTGAAAA ATGCTCTTGT TACTACTGAA    2890

TTCTAGAATC GATAAGCTTC GACCGATGCC CTTGAGAGCC TTCAACCCAG TCAGCTCCTT    2950

CCGGTGGGCG CGGGGCATGA CTATCGTCGC CGCACTTATG ACTGTCTTCT TTATCATGCA    3010

ACTCGTAGGA CAGGTGCCGG CAGCGCTCTG GGTCATTTTC GGCGAGGACC GCTTTCGCTG    3070

GAGCGCGACG ATGATCGGCC TGTCGCTTGC GGTATTCGGA ATCTTGCACG CCCTCGCTCA    3130

AGCCTTCGTC ACTGGTCCCG CCACCAAACG TTTCGGCGAG AAGCAGGCCA TTATCGCCGG    3190

CATGGCGGCC GACGCGCTGG GCTACGTCTT GCTGGCGTTC GTCCAGTAAT GACCTCAGAA    3250

CTCCATCTGG ATTTGTTCAG AACGCTCGGT TGCCGCCGGG CGTTTTTTAT TGGTGAGAAT    3310
```

CGCAGCAACT TGTCGCGCCA ATCGAGCCAT GTCGTCGTCA ACGACCCCCC ATTCAAGAAC 3370
AGCAAGCAGC ATTGAGAACT TTGGAATCCA GTCCCTCTTC CACCTGCTGA GACGCGAGGC 3430
TGGATGGCCT TCCCCATTAT GATTCTTCTC GCTTCCGGCG GCATCGGGAT GCCCGCGTTG 3490
CAGGCCATGC TGTCCAGGCA GGTAGATGAC GACCATCAGG GACAGCTTCA AGGATCGCTC 3550
GCGGCTCTTA CCAGCCTAAC TTCGATCACT GGACCGCTGA TCGTCACGGC GATTTATGCC 3610
GCCTCGGCGA GCACATGGAA CGGGTTGGCA TGGATTGTAG GCGCCGCCCT ATACCTTGTC 3670
TGCCTCCCCG CGTTGCGTCG CGGTGCATGG AGCCGGGCCA CCTCGACCTG AATGGAAGCC 3730
GGCGGCACCT CGCTAACGGA TTCACCACTC CAAGAATTGG AGCCAATCAA TTCTTGCGGA 3790
GAACTGTGAA TGCGCAAACC AACCCTTGGC AGAACATATC CATCGCGTCC GCCATCTCCA 3850
GCAGCCGCAC GCGGCGCATC TCGGGCAGCG TTGGGTCCTG GCCACGGGTG CGCATGATCG 3910
TGCTCCTGTC GTTGAGGACC CGGCTAGGCT GGCGGGGTTG CCTTACTGGT TAGCAGAATG 3970
AATCACCGAT ACGCGAGCGA ACGTGAAGCG ACTGCTGCTG CAAAACGTCT GCGACCTGAG 4030
CAACAACATG AATGGTCTTC GGTTTCCGTG TTTCGTAAAG TCTGGAAACG CGGAAGTCAG 4090
CGCCCTGCAC CATTATGTTC CGGATCTGCA TCGCAGGATG CTGCTGGCTA CCCTGTGGAA 4150
CACCTACATC TGTATTAACG AAGCGCTGGC ATTGACCCTG AGTGATTTTT CTCTGGTCCC 4210
GCCGCATCCA TACCGCCAGT TGTTTACCCT CACAACGTTC CAGTAACCGG GCATGTTCAT 4270
CATCAGTAAC CCGTATCGTG AGCATCCTCT CTCGTTTCAT CGGTATCATT ACCCCCATGA 4330
ACAGAAATTC CCCCTTACAC GGAGGCATCA AGTGACCAAA CAGGAAAAAA CCGCCCTTAA 4390
CATGGCCCGC TTTATCAGAA GCCAGACATT AACGCTTCTG GAGAAACTCA ACGAGCTGGA 4450
CGCGGATGAA CAGGCAGACA TCTGTGAATC GCTTCACGAC CACGCTGATG AGCTTTACCG 4510
CAGCTGCCTC GCGCGTTTCG GTGATGACGG TGAAAACCTC TGACACATGC AGCTCCCGGA 4570
GACGGTCACA GCTTGTCTGT AAGCGGATGC CGGGAGCAGA CAAGCCCGTC AGGGCGCGTC 4630
AGCGGGTGTT GGCGGGTGTC GGGGCGCAGC CATGACCCAG TCACGTAGCG ATAGCGGAGT 4690
GTATACTGGC TTAACTATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA CCATATGCGG 4750
TGTGAAATAC CGCACAGATG CGTAAGGAGA AAATACCGCA TCAGGCGCTC TTCCGCTTCC 4810
TCGCTCACTG ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC GAGCGGTATC AGCTCACTCA 4870
AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG CAGGAAAGAA CATGTGAGCA 4930

AAAGGCCAGC AAAAGGCCAG GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTTCCATAGG 4990
CTCCGCCCCC CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG 5050
ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT 5110
CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC CTTCGGGAAG CGTGGCGCTT 5170
TCTCAATGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC 5230
TGTGTGCACG AACCCCCCGT TCAGCCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT 5290
GAGTCCAACC CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT 5350
AGCAGAGCGA GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC TAACTACGGC 5410
TACACTAGAA GGACAGTATT TGGTATCTGC GCTCTGCTGA AGCCAGTTAC CTTCGGAAAA 5470
AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTTGTT 5530
TGCAAGCAGC AGATTACGCG CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT 5590
ACGGGGTCTG ACGCTCAGTG GAACGAAAAC TCACGTTAAG GGATTTTGGT CATGAGATTA 5650
TCAAAAAGGA TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA ATCAATCTAA 5710
AGTATATATG AGTAAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA GGCACCTATC 5770
TCAGCGATCT GTCTATTTCG TTCATCCATA GTTGCCTGAC TCCCCGTCGT GTAGATAACT 5830
ACGATACGGG AGGGCTTACC ATCTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC 5890
TCACCGGCTC CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT 5950
GGTCCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA 6010
AGTAGTTCGC CAGTTAATAG TTTGCGCAAC GTTGTTGCCA TTGCTGCAGG CATCGTGGTG 6070
TCACGCTCGT CGTTTGGTAT GGCTTCATTC AGCTCCGGTT CCCAACGATC AAGGCGAGTT 6130
ACATGATCCC CCATGTTGTG CAAAAAAGCG GTTAGCTCCT TCGGTCCTCC GATCGTTGTC 6190
AGAAGTAAGT TGGCCGCAGT GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT 6250
ACTGTCATGC CATCCGTAAG ATGCTTTTCT GTGACTGGTG AGTAGCTTCA CGCTGCCGCA 6310
AGCACTCAGG GCGCAAGGGC TGCTAAAGGA AGCGGAACAC GTAGAAAGCC AGTCCGCAGA 6370
AACGGTGCTG ACCCCGGATG AATGTCAGCT ACTGGGCTAT CTGGACAAGG GAAAACGCAA 6430
GCGCAAAGAG AAAGCAGGTA GCTTGCAGTG GGCTTACATG GCGATAGCTA GACTGGGCGG 6490
TTTTATGGAC AGCAAGCGAA CCGGAATTGC CAGCTGGGGC GCCCTCTGGT AAGGTTGGGA 6550
AGCCCTGCAA AGTAAACTGG ATGGCTTTCT TGCCGCCAAG GATCTGATGG CGCAGGGGAT 6610

CAAGATCTGA TCAAGAGACA GGATGAGGAT CGTTTCGCAT GATTGAACAA GATGGATTGC 6670

ACGCAGGTTC TCCGGCCGCT TGGGTGGAGA GGCTATTCGG CTATGACTGG GCACAACAGA 6730

CAATCGGCTG CTCTGATGCC GCCGTGTTCC GGCTGTCAGC GCAGGGGCGC CCGGTTCTTT 6790

TTGTCAAGAC CGACCTGTCC GGTGCCCTGA ATGAACTGCA GGACGAGGCA GCGCGGCTAT 6850

CGTGGCTGGC CACGACGGGC GTTCCTTGCG CAGCTGTGCT CGACGTTGTC ACTGAAGCGG 6910

GAAGGGACTG GCTGCTATTG GGCGAAGTGC CGGGGCAGGA TCTCCTGTCA TCTCACCTTG 6970

CTCCTGCCGA GAAAGTATCC ATCATGGCTG ATGCAATGCG GCGGCTGCAT ACGCTTGATC 7030

CGGCTACCTG CCCATTCGAC CACCAAGCGA AACATCGCAT CGAGCGAGCA CGTACTCGGA 7090

TGGAAGCCGG TCTTGTCGAT CAGGATGATC TGGACGAAGA GCATCAGGGG CTCGCGCCAG 7150

CCGAACTGTT CGCCAGGCTC AAGGCGCGCA TGCCCGACGG CGAGGATCTC GTCGTGACTC 7210

ATGGCGATGC CTGCTTGCCG AATATCATGG TGGAAAATGG CCGCTTTTCT GGATTCATCG 7270

ACTGTGGCCG GCTGGGTGTG GCGGACCGCT ATCAGGACAT AGCGTTGGCT ACCCGTGATA 7330

TTGCTGAAGA GCTTGGCGGC GAATGGGCTG ACCGCTTCCT CGTGCTTTAC GGTATCGCCG 7390

CTCCCGATTC GCAGCGCATC GCCTTCTATC GCCTTCTTGA CGAGTTCTTC TGAGCGGGAC 7450

TCTGGGGTTC GAAATGACCG ACCAAGCGAC GCCCAACCTG CCATCACGAG ATTTCGATTC 7510

CACCGCCGCC TTCTATGAAA GGTTGGGCTT CGGAATCGTT TTCCGGGACG CCGGCTGGAT 7570

GATCCTCCAG CGCGGGGATC TCATGCTGGA GTTCTTCGCC CACCCC 7616

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 304 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
1 5 10 15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
20 25 30

```
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gly Phe Phe Gly Ala Ile Ala Gly Phe Leu Glu Gly Gly Trp Glu
                85                  90                  95

Gly Leu Ile Ala Gly Trp His Gly Tyr Thr Ser His Gly Ala His Gly
            100                 105                 110

Val Ala Val Ala Ala Asp Leu Lys Ser Thr Gln Glu Ala Ile Asn Lys
        115                 120                 125

Ile Thr Lys Asn Leu Asn Tyr Leu Ser Glu Leu Glu Val Lys Asn Leu
    130                 135                 140

Gln Arg Leu Ser Gly Ala Met Asn Glu Leu His Asp Glu Ile Leu Glu
145                 150                 155                 160

Leu Asp Glu Lys Val Asp Asp Leu Arg Ala Asp Thr Ile Ser Ser Gln
                165                 170                 175

Ile Glu Leu Ala Val Leu Leu Ser Asn Glu Gly Ile Ile Asn Ser Glu
            180                 185                 190

Asp Glu His Leu Leu Ala Leu Glu Arg Lys Leu Lys Lys Met Leu Gly
        195                 200                 205

Pro Ser Ala Val Glu Ile Gly Asn Gly Cys Phe Glu Thr Lys His Lys
    210                 215                 220

Cys Asn Gln Thr Cys Leu Asp Arg Ile Ala Ala Gly Thr Phe Asn Ala
225                 230                 235                 240

Gly Asp Phe Ser Leu Pro Thr Phe Asp Ser Leu Asn Ile Thr Ala Ala
                245                 250                 255

Ser Leu Asn Asp Asp Gly Leu Asp Asn His Thr Ile Leu Leu Tyr Tyr
            260                 265                 270

Ser Thr Ala Ala Ser Ser Leu Ala Val Thr Leu Met Ile Ala Ile Phe
        275                 280                 285

Ile Val Tyr Met Val Ser Arg Asp Asn Val Ser Cys Ser Ile Cys Leu
    290                 295                 300
```

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 915 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..912

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG      48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC      96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC     144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45

ACC CTC GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA     192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC     240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80

ATG GGT TTC TTC GGA GCT ATT GCT GGT TTC TTG GAA GGA GGA TGG GAA     288
Met Gly Phe Phe Gly Ala Ile Ala Gly Phe Leu Glu Gly Gly Trp Glu
                 85                  90                  95

GGA ATG ATT GCA GGT TGG CAC GGA TAC ACA TCT CAT GGA GCA CAT GGA     336
Gly Met Ile Ala Gly Trp His Gly Tyr Thr Ser His Gly Ala His Gly
            100                 105                 110

GTG GCA GTG GCA GCA GAC CTT AAG AGT ACA CAA GAA GCT ATA AAC AAG     384
Val Ala Val Ala Ala Asp Leu Lys Ser Thr Gln Glu Ala Ile Asn Lys
        115                 120                 125

ATA ACA AAA AAT CTC AAC TAT TTA AGT GAG CTA GAA GTA AAA AAC CTT     432
Ile Thr Lys Asn Leu Asn Tyr Leu Ser Glu Leu Glu Val Lys Asn Leu
    130                 135                 140
```

```
CAA AGA CTA AGC GGA GCA ATG AAT GAG CTT CAC GAC GAA ATA CTC GAG       480
Gln Arg Leu Ser Gly Ala Met Asn Glu Leu His Asp Glu Ile Leu Glu
145                 150                 155                 160

CTA GAC GAA AAA GTG GAT GAT CTA AGA GCT GAT ACA ATA AGC TCA CAA       528
Leu Asp Glu Lys Val Asp Asp Leu Arg Ala Asp Thr Ile Ser Ser Gln
                165                 170                 175

ATA GAG CTT GCA GTC TTG CTT TCC AAC GAA GGG ATA ATA AAC AGT GAA       576
Ile Glu Leu Ala Val Leu Leu Ser Asn Glu Gly Ile Ile Asn Ser Glu
            180                 185                 190

GAT GAG CAT CTC TTG GCA CTT GAA AGA AAA CTG AAG AAA ATG CTT GGC       624
Asp Glu His Leu Leu Ala Leu Glu Arg Lys Leu Lys Lys Met Leu Gly
        195                 200                 205

CCC TCT GCT GTA GAA ATA GGG AAT GGG TGC TTT GAA ACC AAA CAC AAA       672
Pro Ser Ala Val Glu Ile Gly Asn Gly Cys Phe Glu Thr Lys His Lys
    210                 215                 220

TGC AAC CAG ACT TGC CTA GAC AGG ATA GCT GCT GGC ACC TTT AAT GCA       720
Cys Asn Gln Thr Cys Leu Asp Arg Ile Ala Ala Gly Thr Phe Asn Ala
225                 230                 235                 240

GGA GAT TTT TCT CTT CCC ACT TTT GAT TCA TTA AAC ATT ACT GCT GCA       768
Gly Asp Phe Ser Leu Pro Thr Phe Asp Ser Leu Asn Ile Thr Ala Ala
                245                 250                 255

TCT TTA AAT GAT GAT GGC TTG GAT AAT CAT ACT ATA CTG CTC TAC TAC       816
Ser Leu Asn Asp Asp Gly Leu Asp Asn His Thr Ile Leu Leu Tyr Tyr
            260                 265                 270

TCA ACT GCT GCT TCT AGC TTG GCT GTA ACA TTA ATG ATA GCT ATC TTC       864
Ser Thr Ala Ala Ser Ser Leu Ala Val Thr Leu Met Ile Ala Ile Phe
        275                 280                 285

ATT GTC TAC ATG GTC TCC AGA GAC AAT GTT TCT TGT TCC ATC TGT CTG       912
Ile Val Tyr Met Val Ser Arg Asp Asn Val Ser Cys Ser Ile Cys Leu
    290                 295                 300

TGA                                                                   915
```

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 304 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Gly Phe Phe Gly Ala Ile Ala Gly Phe Leu Glu Gly Gly Trp Glu
                85                  90                  95

Gly Met Ile Ala Gly Trp His Gly Tyr Thr Ser His Gly Ala His Gly
            100                 105                 110

Val Ala Val Ala Ala Asp Leu Lys Ser Thr Gln Glu Ala Ile Asn Lys
        115                 120                 125

Ile Thr Lys Asn Leu Asn Tyr Leu Ser Glu Leu Glu Val Lys Asn Leu
    130                 135                 140

Gln Arg Leu Ser Gly Ala Met Asn Glu Leu His Asp Glu Ile Leu Glu
145                 150                 155                 160

Leu Asp Glu Lys Val Asp Asp Leu Arg Ala Asp Thr Ile Ser Ser Gln
                165                 170                 175

Ile Glu Leu Ala Val Leu Leu Ser Asn Glu Gly Ile Ile Asn Ser Glu
            180                 185                 190

Asp Glu His Leu Leu Ala Leu Glu Arg Lys Leu Lys Lys Met Leu Gly
        195                 200                 205

Pro Ser Ala Val Glu Ile Gly Asn Gly Cys Phe Glu Thr Lys His Lys
    210                 215                 220

Cys Asn Gln Thr Cys Leu Asp Arg Ile Ala Ala Gly Thr Phe Asn Ala
225                 230                 235                 240

Gly Asp Phe Ser Leu Pro Thr Phe Asp Ser Leu Asn Ile Thr Ala Ala
                245                 250                 255

Ser Leu Asn Asp Asp Gly Leu Asp Asn His Thr Ile Leu Leu Tyr Tyr
            260                 265                 270

Ser Thr Ala Ala Ser Ser Leu Ala Val Thr Leu Met Ile Ala Ile Phe
        275                 280                 285
```

Ile Val Tyr Met Val Ser Arg Asp Asn Val Ser Cys Ser Ile Cys Leu
290 295 300

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 918 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

```
ATGGATCCAA ACACTGTGTC AAGCTTTCAG GTAGATTGCT TTCTTTGGCA TGTCCGCAAA     60
CGAGTTGCAG ACCAAGAACT AGGTGATGCC CCATTCCTTG ATCGGCTTCG CCGAGATCAG    120
AAATCCCTAA GAGGAAGGGG CAGCACTCTT GGTCTGGACA TCGAGACAGC CACACGTGCT    180
GGAAAGCAGA TAGTGGAGCG GATTCTGAAA GAAGAATCCG ATGAGGCACT TAAAATGACC    240
ATGGGCGCCC ATATGGGCAT ATTCGGCGCA ATAGCAGGTT TCATAGAAAA TGGTTGGGAG    300
GGAATGATAG ACGGTTGGTA CGGTTTCAGG CATCAAAATT CTGAGGGCAC AGGACAAGCA    360
GCAGATCTTA AAAGCACTCA AGCAGCCATC GACCAAATCA ATGGGAAACT GAATAGGGTA    420
ATCGAGAAGA CGAACGAGAA ATTCCATCAA ATCGAAAAGG AATTCTCAGA AGTAGAAGGG    480
AGAATTCAGG ACCTCGAGAA ATACGTTGAA GACACTAAAA TAGATCTCTG GTCTTACAAT    540
GCGGAGCTTC TTGTCGCTCT GGAGAACCAA CATACAATTG ATCTGACTGA CTCGGAAATG    600
AACAAACTGT TTGAAAAAAC ACGTCGTCAA CTGCGTGAAA ATGCTGAGGA CATGGGCAAT    660
GGTTGCTTCA AAATATACCA CAAATGTGAC AATGCTTGCA TAGGGTCAAT CAGAAATGGG    720
ACTTATGACC ATGATGTATA CAGAGACGAA GCATTAAACA ACCGGTTTCA GATCAAAGGT    780
GTTGAACTGA AGTCAGGATA CAAAGACTGG ATCCTGTGGA TTTCCTTTGC CATATCATGC    840
TTTTTGCTTT GTGTTGTTTT GCTGGGGTTC ATCATGTGGG CCTGCCAAAA AGGCAACATT    900
AGGTGCAACA TTTGCATT                                                  918
```

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

```
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
 1               5                  10                  15

Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
            20                  25                  30

Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
        35                  40                  45

Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
    50                  55                  60

Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
65                  70                  75                  80

Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                85                  90                  95

Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala
            180                 185                 190

Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
        195                 200                 205

Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

```
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu Gly
  1               5                  10                  15

Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly Thr
             20                  25                  30

Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln Ile
         35                  40                  45

Asn Gly Lys Leu Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His
     50                  55                  60

Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu
 65                  70                  75                  80

Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                 85                  90                  95

Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr Asp
            100                 105                 110

Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu
        115                 120                 125

Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys Cys
    130                 135                 140

Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His Asp
145                 150                 155                 160

Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly Val
                165                 170                 175

Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala
            180                 185                 190

Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
        195                 200                 205
```

Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
210 215 220

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Arg Arg Xaa Xaa Arg
1 5

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CGNCGNNNNN NNCGN 15

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

AGRAGRNNNN NNAGR 15

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

CATGGATCAT ATGTTAACAG ATATCAAGGC CTGACTGACT GAGAGCT 47

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

CTCAGTCAGT CAGGCCTTGA TATCTGTTAA CATATGATC 39

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

CATGGGCGCC CATATGGGCA TATTCGGCG 29

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CCGAATATGC CCATATGGGC GCC 23

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

AAACTGTTTG AAAAAACACG TCGTCAACTG CGTGAAAATG CTGACGACAT GGGC 54

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 145 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

```
Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp
1               5                   10                  15

Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile
            20                  25                  30

Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg
        35                  40                  45
```

Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile
50 55 60

Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr
65 70 75 80

Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln
85 90 95

Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp
100 105 110

Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly
115 120 125

Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys
130 135 140

Ile
145

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 145 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp
1 5 10 15

Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile
20 25 30

Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg
35 40 45

Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile
50 55 60

Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr
65 70 75 80

Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln
85 90 95

```
Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp
            100                 105                 110

Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys Val Val Leu Leu Gly
        115                 120                 125

Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys
    130                 135                 140

Ile
145
```

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 690 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..690

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT TGG      48
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
  1               5                  10                  15

CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC CCA TTC      96
His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
             20                  25                  30

CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA AGG GGC AGC     144
Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
         35                  40                  45

ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT GGA AAG CAG ATA     192
Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
     50                  55                  60

GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC     240
Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
 65                  70                  75                  80
```

```
ATG GAT CAT ATG TTA ATT CAG GAC CTC GAG AAA TAC GTT GAA GAC ACT     288
Met Asp His Met Leu Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr
                 85                  90                  95

AAA ATA GAT CTC TGG TCT TAC AAT GCG GAG CTT CTT GTC GCT CTG GAG     336
Lys Ile Asp Leu Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu
            100                 105                 110

AAC CAA CAT ACA ATT GAT CTG ACT GAC TCG GAA ATG AAC AAA CTG TTT     384
Asn Gln His Thr Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe
        115                 120                 125

GAA AAA ACA AGG AGG CAA CTG AGG GAA AAT GCT GAG GAC ATG GGC AAT     432
Glu Lys Thr Arg Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn
    130                 135                 140

GGT TGC TTC AAA ATA TAC CAC AAA TGT GAC AAT GCT TGC ATA GGG TCA     480
Gly Cys Phe Lys Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser
145                 150                 155                 160

ATC AGA AAT GGG ACT TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA     528
Ile Arg Asn Gly Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu
                165                 170                 175

AAC AAC CGG TTT CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA TAC AAA     576
Asn Asn Arg Phe Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys
            180                 185                 190

GAC TGG ATC CTG TGG ATT TCC TTT GCC ATA TCA TGC TTT TTG CTT TGT     624
Asp Trp Ile Leu Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys
        195                 200                 205

GTT GTT TTG CTG GGG TTC ATC ATG TGG GCC TGC CAA AAA GGC AAC ATT     672
Val Val Leu Leu Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile
    210                 215                 220

AGG TGC AAC ATT TGC ATT                                             690
Arg Cys Asn Ile Cys Ile
225                 230
```

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 230 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

```
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu Trp
 1               5                  10                  15

His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala Pro Phe
            20                  25                  30

Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly Arg Gly Ser
        35                  40                  45

Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala Gly Lys Gln Ile
    50                  55                  60

Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu Ala Leu Lys Met Thr
65                  70                  75                  80

Met Asp His Met Leu Ile Gln Asp Leu Glu Lys Tyr Val Glu Asp Thr
                85                  90                  95

Lys Ile Asp Leu Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu
            100                 105                 110

Asn Gln His Thr Ile Asp Leu Thr Asp Ser Glu Met Asn Lys Leu Phe
        115                 120                 125

Glu Lys Thr Arg Arg Gln Leu Arg Glu Asn Ala Glu Asp Met Gly Asn
    130                 135                 140

Gly Cys Phe Lys Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser
145                 150                 155                 160

Ile Arg Asn Gly Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu
                165                 170                 175

Asn Asn Arg Phe Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys
            180                 185                 190

Asp Trp Ile Leu Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys
        195                 200                 205

Val Val Leu Leu Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile
    210                 215                 220
```

```
Arg Cys Asn Ile Cys Ile
225                 230
```

WHAT IS CLAIMED IS:

1. A vaccine for stimulating protection in animals against infection by influenza virus which comprises an effective amount of an immunogenic fragment of the HA2 subunit of an HA protein selected from the group consisting of a Type A subtype influenza virus and a Type B influenza virus.

2. The vaccine according to claim 1 wherein said Type A subunit is H3N2.

3. The vaccine according to claim 1 wherein the polypeptide is fused to a second polypeptide.

4. The vaccine according to claim 3 wherein the second polypeptide comprises the N terminal amino acids of influenza NS1 protein.

5. The vaccine according to claim 1 wherein the immunogenic fragment of the HA2 subunit is selected from the group consisting of a peptide comprising amino acids 1 to 221 of the H3HA2 subtype, a peptide comprising amino acids 77 to 221 of the H3HA2 subtype, a peptide comprising amino acids 1 to 223 of the BHA2 Type, and a peptide comprising amino acids 41 to 223 of the BHA2 type.

6. The vaccine according to claim 5 comprising $NS1_{(1-81)}H3HA2_{(1-221)}$ SEQ ID NO: 10.

7. The vaccine according to claim 5 comprising $NS1_{(1-81)}H3HA2_{(77-221)}$ SEQ ID NO: 12.

8. The vaccine according to claim 5 comprising $NS1_{(1-42)}BLHA2_{(41-223)}$ SEQ ID NO: 14.

9. The vaccine according to claim 5 comprising $NS1_{(1-81)}BLHA2_{(1-223)}$ SEQ ID NO: 57.

10. The vaccine according to claim 5 comprising NS1$_{(1-81)}$H3HA2$_{(1-221)}$ SEQ ID NO:10 and NS1$_{(1-81)}$BLHA2$_{(1-223)}$(met-leu) SEQ ID NO: 55.

11. A protein comprising an immunogenic fragment of the HA2 subunit of an HA protein selected from the group consisting of Type A subtype or Type B influenza virus.

12. The protein according to claim 11 wherein said Type A subtype is H3N2.

13. The protein according to claim 11 wherein the peptide containing the immunogenic fragment is fused to a second peptide or protein.

14. The protein according to claim 13 wherein the second peptide comprises the N terminal amino acids of a NS1 protein.

15. The protein according to claim 11 wherein the immunogenic fragment of the HA2 subunit is selected from the group consisting of a peptide comprising amino acids 1 to 221 of the H3HA2 subunit, a peptide comprising amino acids 77 to 221 of the H3HA2 subunit, a peptide comprising amino acids 1-223 of the BHA2 subunit, and a peptide comprising amino acids 41-223 of the BHA2 subunit.

16. A polypeptide NS1$_{(1-81)}$H3HA2$_{(1-221)}$ SEQ ID NO: 10.

17. A polypeptide NS1$_{(1-81)}$H3HA2$_{(77-221)}$ SEQ ID NO: 12.

18. A polypeptide NS1$_{(1-41)}$BLHA2$_{(41-223)}$ SEQ ID NO: 14.

19. A polypeptide NS1$_{(1-81)}$BLHA2$_{(1-223)}$ SEQ ID NO: 57.

20. A polypeptide NS1$_{(1-81)}$BLHA2$_{(1-223)}$(met-leu) SEQ ID NO: 55.

21. A DNA molecule comprising a coding sequence for an immunogenic fragment of the HA2 subunit of an HA protein selected from the group consisting of a Type A subtype or Type B influenza virus.

22. The DNA molecule according to claim 21 wherein said Type A subunit is H3N2.

23. The DNA molecule according to claim 22 comprising a coding sequence for the polypeptide $NS1_{(1-81)}H3HA2_{(1-221)}$ SEQ ID NO: 10.

24. The DNA molecule according to claim 21 comprising a coding sequence for the polypeptide $NS1_{(1-42)}BLHA2_{(41-223)}$ SEQ ID NO: 14.

25. The DNA molecule according to claim 21 comprising a coding sequence for the polypeptide $NS1_{(1-81)}H3HA2_{(77-221)}$ SEQ ID NO: 12.

26. The DNA molecule according to claim 21 comprising a coding sequence for the polypeptide $NS1_{(1-81)}BLHA2_{(1-223)}$ SEQ ID NO: 57.

27. A vector pOTS208NS1BLmut2 SEQ ID NO: 54.

28. A microorganism transformed with a DNA molecule comprising a coding sequence for an immunogenic fragment of the HA2 subunit of an HA protein selected from the group consisting of a Type A subtype or Type B influenza virus.

29. The microorganism according to claim 28 wherein said Type A subunit is H3N2.

30. The microorganism according to claim 28 wherein said DNA molecule comprises a coding sequence for the polypeptide $NS1_{(1-81)}H3HA2_{(1-221)}$ SEQ ID NO: 10.

31. The microorganism according to claim 28 wherein said DNA molecule comprises a coding sequence for the polypeptide $NS1_{(1-81)}BLHA2_{(1-223)}$ SEQ ID NO: 57.

32. The microorganism according to claim 28 wherein said DNA molecule comprises a coding sequence for the polypeptide $NS1_{(1-81)}BLHA2_{(1-223)}$(met-leu) SEQ ID NO: 55.

33. A combination vaccine for stimulating protection in animals against infection by influenza virus which comprises a first polypeptide having an immunogenic fragment of the HA2 subunit of an influenza H3 subtype virus and a second polypeptide selected from the group consisting of a polypeptide having an immunogenic fragment of the HA2 subunit of a Type B influenza virus, and a polypeptide having an immunogenic fragment of the HA2 subunit of an H1 subtype influenza virus, and a polypeptide having an immunogenic fragment of the HA2 subunit of an H2 subtype influenza virus.

34. The combination vaccine according to claim 33 wherein the first polypeptide is selected from the group consisting of $NS1_{(1-81)}H3HA2_{(1-221)}$ SEQ ID NO: 10 and $NS1_{(1-81)}H3HA2_{(77-221)}$ SEQ ID NO: 12.

35. The combination vaccine according to claim 33 wherein the second polypeptide is a polypeptide having an immunogenic fragment of the HA2 subunit of an H1 subtype influenza virus.

36. The combination vaccine according to claim 33 wherein said second polypeptide is selected from the group consisting of C13 SEQ ID NO: 16, D SEQ ID NO: 18, C13 short SEQ ID NO: 20, D short SEQ ID NO: 22, A SEQ ID NO: 24, C SEQ ID NO: 26, ΔD SEQ ID NO: 27, Δ13 SEQ ID NO: 28, M SEQ ID NO: 29, ΔM SEQ ID NO: 30, ΔM+ SEQ ID NO: 32, and $H1HA2_{66-222}$ SEQ ID NO: 34.

37. The combination vaccine according to claim 33 wherein said second polypeptide is $NS1_{(1-42)}BLHA2_{(41-223)}$ SEQ ID NO: 14.

38. The combination vaccine according to claim 33 wherein said second polypeptide is $NS1_{(1-81)}BLHA2_{(1-223)}$ SEQ ID NO: 57.

39. A combination vaccine for stimulating protection in animals against infection by influenza virus which comprises a first polypeptide having an immunogenic fragment of the HA2 subunit of an influenza H3 subtype virus, a

second polypeptide having an immunogenic fragment of the HA2 subunit of an influenza B Type virus, and a third polypeptide selected from the group consisting of a polypeptide having an immunogenic fragment of the HA2 subunit of an H1 subtype influenza virus and a polypeptide having an immunogenic fragment of the HA2 subunit of an H2 subtype influenza virus.

40. The combination vaccine according to claim 39 wherein the first polypeptides is $NS1_{(1-81)}H3HA2_{(1-221)}$ SEQ ID NO: 10, the second polypeptide is $NS1_{(1-81)}BHA2_{(1-223)}$(met-leu) SEQ ID NO: 57, and the third polypeptide is $NS1_{(1-81)}HA2_{(65-222)}$ SEQ ID NO: 18.

FIGURE 1

```
(a)  ---------- ---------- ---------- ---------- ----------
(b)  ---------- ---------- ---------- ---------- ----------
(c)  --tc----t- -a--c--t-- c-----t--t ---ggg--a- --act-----
(d)  GGCATATTCG GCGCAATAGC AGGTTTCATA GAAAATGGTT GGGAGGGAAT    50

(a)  ---------- ---------- ---------- ------t--- ----------
(b)  ---------- ---------- ---------- ------c--- ----------
(c)  ------t--a ---------- atcat----- g---gaac-- --at----ct
(d)  GATAGACGGT TGGTACGGTT TCAGGCATCA AAATTC-GAG GGCACAGGAC    100

(a)  ---------- ---------- ---------- ---------- ----------
(b)  ---------- ---------- ---------- ---------- ----------
(c)  -t-----g-- --aa------ --a---aat- ----ta--gg g--t-caaac
(d)  AAGCAGCAGA TCTTAAAAGC ACTCAAGCAG CCATCGACCA AATCAATGGG    150

(a)  ---------- ---------- ---------- ---------- ----------
(b)  ---------- ---------- ---------- ---------- ----------
(c)  --gg----ct ct--t----- ---a-t---- attc-----a cagctg-g-g
(d)  AAACTGAATA GGGTAATCGA GAAGACGAAC GAGAAATTCC ATCAAATCGA    200

(a)  ---------- ---------- ---------- ---------- ----------
(b)  ---------- ---------- ---------- ---------- ----------
(c)  t--a------ aaca--t--- --aaa--g-- gg-aa-tt-a a-t---a-a-
(d)  AAAGGAATTC TCAGAAGTAG AAGGGAGAAT TCAGGACCTC GAGAAATACG    250

(a)  ---------- ---------- ---------- ---------- ----------
(b)  ---------- ---------- ---------- ---------- ----------
(c)  ----t--tgg atttc-g--c a-t---a-a- -t-----a-- at-gt-a--t
(d)  TTGAAGACAC TAAAATAGAT CTCTGGTCTT ACAATGCGGA GCTTCTTGTC    300

(a)  ---------- ---------- ---------- ---------- ----------
(b)  ---------- ---------- ---------- ---------- ----------
(c)  cta-----a- -tg--agg-- tc-g---t-c ca------aa -tg----g--
(d)  GCTCTGGAGA ACCAACATAC AATTGATCTG ACTGACTCGG AAATGAACAA    350

(a)  ---------- ---------- ---------- ---------- ----------
(b)  ---------- ---------- ---------- ---------- ----------
(c)  t----a---g ---gt--aa- -c---t-a-a -a-t-----c a-a--a--c-
(d)  ACTGTTTGAA AAAACAAGGA GGCAACTGAG GGAAAATGCT GAGGACATGG    400

(a)  ---------- ---------- ---------- ---------- ----------
(b)  ---------- ---------- ---------- ---------- ----------
(c)  -a-----a-- t--tg-gt-c --------g- ---------a a-----g-aa
(d)  GCAATGGTTG CTTCAAAATA TACCACAAAT GTGACAATGC TTGCATAGGG    450

(a)  ---------- ---------- ---------- ---------- ----------
(b)  ---------- ---------- ---------- ---------- ----------
(c)  agtg-a---- ---------- ---tt--ccc aa---ttc-- -a--gt--aa
(d)  TCAATCAGAA ATGGGACTTA TGACCATGAT GTATACAGAG ACGAAGCATT    500

(a)  ---------- ---------- ---------- ---------- ----------
(b)  ---------- ---------- ---------- ---------- ----------
(c)  gttg---a-- gaaa--g-ag -t--a--ga- -t--g-a--- atgggg-tct
(d)  AAACAACCGG TTTCAGATC[illegible] AAGGTGTTGA ACTGAAGTCA GGATACAAAG    550
```

FIGURE 1 (cont'd)

```
(a)   ---------- ---------- ---------- ---------- ----------
(b)   ---------- ---------- ---------- ---------- ----------
(c)   -tca---t-- -gc---c-a- -caa-tg-cg -ca-t-cac- -g-gct-t-g
(d)   ACTGGATCCT GTGGATTTCC TTTGCCATAT CATGCTTTTT GCTTTGTGTT    600

(a)   ---------- ---------. ..-------- -----g---- ----------
(b)   ---------- ---------. ..-------- -----a---- ----------
(c)   --c-cc---- --gca----g tt-c---atg --ttct--t- -atctt-gca
(d)   GTTTTGCTGG GGTTCATCA- --TGTGGGCC TGCCA-AAAG GCAACATTAG    650

(a)   ---------- ---------.
(b)   ---------- ---------.
(c)   -----ga--a -----c---g
(d)   GTGCAACATT TGCATTTGA-                                     670
```

FIGURE 2

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT          42
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe
 1               5                  10

CTT TGG CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT          84
Leu Trp His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly
 15                  20                  25

GAT GCC CCA TTC CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC         126
Asp Ala Pro Phe Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser
     30                  35                  40

CTA AGA GGA AGG GGC AGC ACT CTT GGT CTG GAC ATC GAG ACA         168
Leu Arg Gly Arg Gly Ser Thr Leu Gly Leu Asp Ile Glu Thr
         45                  50                  55

GCC ACA CGT GCT GGA AAG CAG ATA GTG GAG CGG ATT CTG AAA         210
Ala Thr Arg Ala Gly Lys Gln Ile Val Glu Arg Ile Leu Lys
             60                  65                  70

GAA GAA TCC GAT GAG GCA CTT AAA ATG ACC ATG GGC GCC CAT         252
Glu Glu Ser Asp Glu Ala Leu Lys Met Thr Met Gly Ala His
                 75                  80

ATG GGC ATA TTC GGC GCA ATA GCA GGT TTC ATA GAA AAT GGT         294
Met Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly
 85                  90                  95

TGG GAG GGA ATG ATA GAC GGT TGG TAC GGT TTC AGG CAT CAA         336
Trp Glu Gly Met Ile Asp Gly Trp Tyr Gly Phe Arg His Gln
    100                 105                 110

AAT TCT GAG GGC ACA GGA CAA GCA GCA GAT CTT AAA AGC ACT         378
Asn Ser Glu Gly Thr Gly Gln Ala Ala Asp Leu Lys Ser Thr
        115                 120                 125

CAA GCA GCC ATC GAC CAA ATC AAT GGG AAA CTG AAT AGG GTA         420
Gln Ala Ala Ile Asp Gln Ile Asn Gly Lys Leu Asn Arg Val
            130                 135                 140

ATC GAG AAG ACG AAC GAG AAA TTC CAT CAA ATC GAA AAG GAA         462
Ile Glu Lys Thr Asn Glu Lys Phe His Gln Ile Glu Lys Glu
                145                 150

TTC TCA GAA GTA GAA GGG AGA ATT CAG GAC CTC GAG AAA TAC         504
Phe Ser Glu Val Glu Gly Arg Ile Gln Asp Leu Glu Lys Tyr
155                 160                 165

GTT GAA GAC ACT AAA ATA GAT CTC TGG TCT TAC AAT GCG GAG         546
Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala Glu
    170                 175                 180
```

FIGURE 2 (cont'd)

```
CTT CTT GTC GCT CTG GAG AAC CAA CAT ACA ATT GAT CTG ACT    588
Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr
        185                 190                 195

                                            C T C T
GAC TCG GAA ATG AAC AAA CTG TTT GAA AAA ACA AGG AGG CAA    630
Asp Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln
            200                 205                 210
    C T
CTG AGG GAA AAT GCT GAG GAC ATG GGC AAT GGT TGC TTC AAA    672
Leu Arg Glu Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys
                215                 220

ATA TAC CAC AAA TGT GAC AAT GCT TGC ATA GGG TCA ATC AGA    714
Ile Tyr His Lys Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg
225                 230                 235

AAT GGG ACT TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA    756
Asn Gly Thr Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu
    240                 245                 250

AAC AAC CGG TTT CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA    798
Asn Asn Arg Phe Gln Ile Lys Gly Val Glu Leu Lys Ser Gly
        255                 260                 265

TAC AAA GAC TGG ATC CTG TGG ATT TCC TTT GCC ATA TCA TGC    840
Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala Ile Ser Cys
            270                 275                 280

TTT TTG CTT TGT GTT GTT TTG CTG GGG TTC ATC ATG TGG GCC    882
Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp Ala
                285                 290

TGC CAA AAA GGC AAC ATT AGG TGC AAC ATT TGC ATT            918
Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
295                 300                 305
```

FIGURE 3

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TGC TTT CTT      45
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Cys Phe Leu
  1               5                  10                  15

TGG CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC      90
Trp His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala
                 20                  25                  30

CCA TTC CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA AGA GGA     135
Pro Phe Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu Arg Gly
                 35                  40                  45

AGG GGC AGC ACT CTT GGT CTG GAC ATC GAG ACA GCC ACA CGT GCT     100
Arg Gly Ser Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg Ala
                 50                  55                  60

GGA AAG CAG ATA GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG     225
Gly Lys Gln Ile Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu
                 65                  70                  75

GCA CTT AAA ATG ACC ATG GAT CAT ATG TTA ATT CAG GAC CTC GAG     270
Ala Leu Lys Met Thr Met Asp His Met Leu Ile Gln Asp Leu Glu
                 80                  85                  90

AAA TAC GTT GAA GAC ACT AAA ATA GAT CTC TGG TCT TAC AAT GCG     315
Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn Ala
                 95                 100                 105

GAG CTT CTT GTC GCT CTG GAG AAC CAA CAT ACA ATT GAT CTG ACT     360
Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr
                110                 115                 120

GAC TCG GAA ATG AAC AAA CTG TTT GAA AAA ACA AGG AGG CAA CTG     405
Asp Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Arg Arg Gln Leu
                125                 130                 135

AGG GAA AAT GCT GAG GAC ATG GGC AAT GGT TGC TTC AAA ATA TAC     450
Arg Glu Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr
                140                 145                 150

CAC AAA TGT GAC AAT GCT TGC ATA GGG TCA ATC AGA AAT GGG ACT     495
His Lys Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr
                155                 160                 165

TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA AAC AAC CGG TTT     540
Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe
                170                 175                 180

CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA TAC AAA GAC TGG ATC     585
Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile
                185                 190                 195
```

FIGURE 3 (cont'd)

```
CTG TGG ATT TCC TTT GCC ATA TCA TGC TTT TTG CTT TGT GTT GTT      630
Leu Trp Ile Ser Phe Ala Ile Ser Cys Phe Leu Leu Cys Val Val
                200                 205                 210

TTG CTG GGG TTC ATC ATG TGG GCC TGC CAA AAA GGC AAC ATT AGG      675
Leu Leu Gly Phe Ile Met Trp Ala Cys Gln Lys Gly Asn Ile Arg
                215                 220                 225

TGC AAC ATT TGC ATT                                              690
Cys Asn Ile Cys Ile
                230
```

FIGURE 4

```
ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA GAT TCC TTT CTT     45
Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val Asp Ser Phe Leu
  1               5                  10                  15

TGG CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA GGT GAT GCC     90
Trp His Val Arg Lys Arg Val Ala Asp Gln Glu Leu Gly Asp Ala
                 20                  25                  30

CCA TTC CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC ATG CAT GGA    135
Pro Phe Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Met His Gly
                 35                  40                  45

TCA TAT GTT AAC AAG ACA CAA GAA GCT ATA AAC AAG ATA ACA AAA    180
Ser Tyr Val Asn Lys Thr Gln Glu Ala Ile Asn Lys Ile Thr Lys
                 50                  55                  60

AAT CTC AAC TAT TTA AGT GAG CTA GAA GTA AAA AAC CTT CAA AGA    225
Asn Leu Asn Tyr Leu Ser Glu Leu Glu Val Lys Asn Leu Gln Arg
                 65                  70                  75

CTA AGC GGA GCA ATG AAT GAG CTT CAC GAC GAA ATA CTC GAG CTA    270
Leu Ser Gly Ala Met Asn Glu Leu His Asp Glu Ile Leu Glu Leu
                 80                  85                  90

GAC GAA AAA GTG GAT GAT CTA AGA GCT GAT ACA ATA AGC TCA CAA    315
Asp Glu Lys Val Asp Asp Leu Arg Ala Asp Thr Ile Ser Ser Gln
                 95                 100                 105

ATA GAG CTT GCA GTC TTG CTT TCC AAC GAA GGG ATA ATA AAC AGT    360
Ile Glu Leu Ala Val Leu Leu Ser Asn Glu Gly Ile Ile Asn Ser
                110                 115                 120

GAA GAT GAG CAT CTC TTG GCA CTT GAA AGA AAA CTG AAG AAA ATG    405
Glu Asp Glu His Leu Leu Ala Leu Glu Arg Lys Leu Lys Lys Met
                125                 130                 135

CTT GGC CCC TCT GCT GTA GAA ATA GGGAAT GGG TGC TTT GAA ACC     450
Leu Gly Pro Ser Ala Val Glu Ile Gly Asn Gly Cys Phe Glu Thr
                140                 145                 150

AAA CAC AAA TGC AAC CAG ACT TGC CTA GAC AGG ATA GCT GCT GGC    495
Lys His Lys Cys Asn Gln Thr Cys Leu Asp Arg Ile Ala Ala Gly
                155                 160                 165

ACC TTT AAT GCA GGA GAT TTT TCT CTT CCC ACT TTT GAT TCA TTA    540
Thr Phe Asn Ala Gly Asp Phe Ser Leu Pro Thr Phe Asp Ser Leu
                170                 175                 180

AAC ATT ACT GCT GCA TCT TTA AAT GAT GAT GGC TTG GAT AAT CAT    585
Asn Ile Thr Ala Ala Ser Leu Asn Asp Asp Gly Leu Asp Asn His
                185                 190                 195
```

FIGURE 4 (cont'd)

```
ACT ATA CTG CTC TAC TAC TCA ACT GCT GCT TCT AGC TTG GCT GTA     630
Thr Ile Leu Leu Tyr Tyr Ser Thr Ala Ala Ser Ser Leu Ala Val
                200                 205                 210

ACA TTA ATG ATA GCT ATC TTC ATT GTC TAC ATG GTC TCC AGA GAC     675
Thr Leu Met Ile Ala Ile Phe Ile Val Tyr Met Val Ser Arg Asp
                215                 220                 225

AAT GTT TCT TGT TCC ATC TGT CTG                                699
Asn Val Ser Cys Ser Ile Cys Leu
                230
```

FIGURE 5

AATTCTCATG TTTGACAGCT TATCATCGAT AAGCTTCAGT TGAAGATATT AAGAACAGCC 60
TCGCAGATGA CGAATCATTG GGATTCCCAT CTTTTTTGTT TGTTGAAGGC GACACCATTG 120
GTTTTGCCAG AACTGTTTTC GGGCCGACCA CATCCGATCT GACAGATTTT TTAATCGGGA 180
AAGGAATGTC ATTAAGCAGT GGAGAGCGCG TTCAGATAGA GCCACTGATG AGGGGAACCA 240
CCAAAGACGA TGTTATGCAT ATGCATTTCA TCGGCCGAAC AACGGTGAAG GTAGAAGCCA 300
AGCTACCTGT ATTTGGCGAT ATATTAAAGG TCTTAGGGGC AACAGATATT GAAGGGGAGC 360
TTTTTGACTC ATTGGATATA GTCATTAAGC CAAAATTTAA AAGGGATATA AAAAAGGTTG 420
CCAAGGATAT TATTTTTAAC CCGTCACCTC AATTTTCAGA CATTAGCCTG CGGGCAAAAG 480
ATGAGGCCGG AGATATTTTA ACAGAACATT ATCTATCAGA AAAAGGCCAT CTCTCAGCGC 540
CTCTGAACAA GGTCACCAAT GCTGAGATAG CTGAAGAGAT GGCATATTGC TACGCAAGAA 600
TGAAAAGTGA TATACTGGAA TGTTTTAAAA GGCAGGTGGG CAAAGTTAAG GATTAATTAT 660
CAGGAGTAAT TATGCGGAAC AGAATCATGC CTGGTGTTTA CATAGTAATA ATTCCTTACG 720
TTATCGTAAG CATTTGCTAT CTCCTTTTCC GCCACTACAT TCCTGGTGTT TCTTTTTCAG 780
CTCATAGAGA TGGTCTTGGG GCGACATTGT CATCATATGC AGGAACCATG ATTGCAATCC 840
TGATTGCTGC CTTGACGTTT CTAATCGGAA GCAGAACGCG CCGACTGGCC AAGATTAGAG 900
AGTATGGGTA TATGACATCG GTAGTTATTG TCTATGCCCT TAGTTTTGTT GAGCTTGGAG 960
CTTTGTTTTT CTGCGGGTTA TTGCTTCTTT CCAGCATAAG CGGCTACATG ATACCCACTA 1020
TCGCCATCGG CATTGCCTCT GCATCGTTCA TTCATATATG CATCCTTGTT TTCCAACTAT 1080
ATAATTTGAC CAGAGAACAA GAATAACCCG GCCTCAGCGC CGGGTTTTCT TTGCCTCACG 1140
ATCGCCCCCA AAACACATAA CCAATTGTAT TTATTGAAAA ATAAATAGAT ACAACTCACT 1200
AAACATAGCA ATTCAGATCT CTCACCTACC AAACAATGCC CCCCTGCAAA AAATAAATTC 1260
ATATAAAAAA CATACAGATA ACCATCTGCG GTGATAAATT ATCTCTGGCG GTGTTGACAT 1320
AAATACCACT GGCGGTGATA CTGAGCACAT CAGCAGGACG CACTGACCAC CATGAAGGTG 1380
ACGCTCTTAA AAATTAAGCC CTGAAGAAGG GCAGCATTCA AAGCAGAAGG CTTTGGGGTG 1440
TGTGATACGA AACGAAGCAT TGGCCGTAAG TGCGATTCCG GATTAGCTGC CAATGTGCCA 1500
ATCGCGGGGG GTTTTCGTTC AGGACTACAA CTGCCACACA CCACCAAAGC TAACTGACAG 1560
GAGAATCCAG ATGGATGCAC AAACACGCCG CCGCGAACGT CGCGCAGAGA AACAGGCTCA 1620

FIGURE 5 (cont'd)

```
ATGGAAAGCA GCAAATCCCC TGTTGGTTGG GGTAAGCGCA AAACCAGTTC CGAAAGATTT 1680

TTTTAACTAT AAACGCTGAT GGAAGCGTTT ATGCGGAAGA GGTAAAGCCC TTCCCGAGTA 1740

ACAAAAAAAC AACAGCATAA ATAACCCCGC TCTTACACAT TCCAGCCCTG AAAAAGGGCA 1800

TCAAATTAAA CCACACCTAT GGTGTATGCA TTTATTTGCA TACATTCAAT CAATTGTTAT 1860

CTAAGGAAAT ACTTACAT ATG GAT CCA AAC ACT GTG TCA AGC TTT CAG GTA    1911
                    Met Asp Pro Asn Thr Val Ser Ser Phe Gln Val
                     1               5                   10

GAT TGC TTT CTT TGG CAT GTC CGC AAA CGA GTT GCA GAC CAA GAA CTA    1959
Asp Cys Phe Leu Trp His Val Arg Lys Arg Val Ala Asp Gln Glu Leu
            15                  20                  25

GGT GAT GCC CCA TTC CTT GAT CGG CTT CGC CGA GAT CAG AAA TCC CTA    2007
Gly Asp Ala Pro Phe Leu Asp Arg Leu Arg Arg Asp Gln Lys Ser Leu
        30                  35                  40

AGA GGA AGG GGC AGC ACC CTC GGT CTG GAC ATC GAG ACA GCC ACA CGT    2055
Arg Gly Arg Gly Ser Thr Leu Gly Leu Asp Ile Glu Thr Ala Thr Arg
    45                  50                  55

GCT GGA AAG CAG ATA GTG GAG CGG ATT CTG AAA GAA GAA TCC GAT GAG    2103
Ala Gly Lys Gln Ile Val Glu Arg Ile Leu Lys Glu Glu Ser Asp Glu
60                  65                  70                  75

GCA CTT AAA ATG ACC ATG GGT TTC TTC GGA GCT ATT GCT GGT TTC TTG    2151
Ala Leu Lys Met Thr Met Gly Phe Phe Gly Ala Ile Ala Gly Phe Leu
                80                  85                  90

      A   A           A  ATG
GAA GGT GGT TGG GAA GGT CTC ATT GCA GGT TGG CAC GGA TAC ACA TCT    2199
Glu Gly Gly Trp Glu Gly Leu Ile Ala Gly Trp His Gly Tyr Thr Ser
            95              Met     100                 105

CAT GGA GCA CAT GGA GTG GCA GTG GCA GCA GAC CTT AAG AGT ACA CAA    2247
His Gly Ala His Gly Val Ala Val Ala Ala Asp Leu Lys Ser Thr Gln
        110                 115                 120

GAA GCT ATA AAC AAG ATA ACA AAA AAT CTC AAC TAT TTA AGT GAG CTA    2295
Glu Ala Ile Asn Lys Ile Thr Lys Asn Leu Asn Tyr Leu Ser Glu Leu
    125                 130                 135

GAA GTA AAA AAC CTT CAA AGA CTA AGC GGA GCA ATG AAT GAG CTT CAC    2343
Glu Val Lys Asn Leu Gln Arg Leu Ser Gly Ala Met Asn Glu Leu His
140                 145                 150                 155

GAC GAA ATA CTC GAG CTA GAC GAA AAA GTG GAT GAT CTA AGA GCT GAT    2391
Asp Glu Ile Leu Glu Leu Asp Glu Lys Val Asp Asp Leu Arg Ala Asp
                160                 165                 170
```

FIGURE 5 (cont'd)

```
                                                              G
ACA ATA AGC TCA CAA ATA GAG CTT GCA GTC TTG CTT TCC AAC GAA GGT    2439
Thr Ile Ser Ser Gln Ile Glu Leu Ala Val Leu Leu Ser Asn Glu Gly
            175                 180                 185

 A   A                   T
ATC ATC AAC AGT GAA GAC GAG CAT CTC TTG GCA CTT GAA AGA AAA CTG    2487
Ile Ile Asn Ser Glu Asp Glu His Leu Leu Ala Leu Glu Arg Lys Leu
        190                 195                 200

                                          C   T   T   G
AAG AAA ATG CTT GGC CCC TCT GCT GTA GAA ATA GGG AAC GGT TGC TTT    2535
Lys Lys Met Leu Gly Pro Ser Ala Val Glu Ile Gly Asn Gly Cys Phe
    205                 210                 215

GAA ACC AAA CAC AAA TGC AAC CAG ACT TGC CTA GAC AGG ATA GCT GCT    2583
Glu Thr Lys His Lys Cys Asn Gln Thr Cys Leu Asp Arg Ile Ala Ala
220                 225                 230                 235

GGC ACC TTT AAT GCA GGA GAT TTT TCT CTT CCC ACT TTT GAT TCA TTA    2631
Gly Thr Phe Asn Ala Gly Asp Phe Ser Leu Pro Thr Phe Asp Ser Leu
            240                 245                 250

AAC ATT ACT GCT GCA TCT TTA AAT GAT GAT GGC TTG GAT AAT CAT ACT    2679
Asn Ile Thr Ala Ala Ser Leu Asn Asp Asp Gly Leu Asp Asn His Thr
            255                 260                 265

ATA CTG CTC TAC TAC TCA ACT GCT GCT TCT AGC TTG GCT GTA ACA TTA    2727
Ile Leu Leu Tyr Tyr Ser Thr Ala Ala Ser Ser Leu Ala Val Thr Leu
        270                 275                 280

ATG ATA GCT ATC TTC ATT GTC TAC ATG GTC TCC AGA GAC AAT GTT TCT    2775
Met Ile Ala Ile Phe Ile Val Tyr Met Val Ser Arg Asp Asn Val Ser
    285                 290                 295

TGT TCC ATC TGT CTG TGAGGGAGAT TAAGCCCTGT GTTTTCCTTT ACTGTAGTGC    2830
Cys Ser Ile Cys Leu
300

TCATTTGCTT GTCACCATTA CAAAGAAACG TTATTGAAAA ATGCTCTTGT TACTACTGAA 2890

TTCTAGAATC GATAAGCTTC GACCGATGCC CTTGAGAGCC TTCAACCCAG TCAGCTCCTT 2950

CCGGTGGGCG CGGGGCATGA CTATCGTCGC CGCACTTATG ACTGTCTTCT TTATCATGCA 3010

ACTCGTAGGA CAGGTGCCGG CAGCGCTCTG GGTCATTTTC GGCGAGGACC GCTTTCGCTG 3070

GAGCGCGACG ATGATCGGCC TGTCGCTTGC GGTATTCGGA ATCTTGCACG CCCTCGCTCA 3130

AGCCTTCGTC ACTGGTCCCG CCACCAAACG TTTCGGCGAG AAGCAGGCCA TTATCGCCGG 3190

CATGGCGGCC GACGCGCTGG GCTACGTCTT GCTGGCGTTC GTCCAGTAAT GACCTCAGAA 3250
```

FIGURE 5 (cont'd)

CTCCATCTGG ATTTGTTCAG AACGCTCGGT TGCCGCCGGG CGTTTTTTAT TGGTGAGAAT 3310
CGCAGCAACT TGTCGCGCCA ATCGAGCCAT GTCGTCGTCA ACGACCCCCC ATTCAAGAAC 3370
AGCAAGCAGC ATTGAGAACT TTGGAATCCA GTCCCTCTTC CACCTGCTGA GACGCGAGGC 3430
TGGATGGCCT TCCCCATTAT GATTCTTCTC GCTTCCGGCG GCATCGGGAT GCCCGCGTTG 3490
CAGGCCATGC TGTCCAGGCA GGTAGATGAC GACCATCAGG GACAGCTTCA AGGATCGCTC 3550
GCGGCTCTTA CCAGCCTAAC TTCGATCACT GGACCGCTGA TCGTCACGGC GATTTATGCC 3610
GCCTCGGCGA GCACATGGAA CGGGTTGGCA TGGATTGTAG GCGCCGCCCT ATACCTTGTC 3670
TGCCTCCCCG CGTTGCGTCG CGGTGCATGG AGCCGGGCCA CCTCGACCTG AATGGAAGCC 3730
GGCGGCACCT CGCTAACGGA TTCACCACTC CAAGAATTGG AGCCAATCAA TTCTTGCGGA 3790
GAACTGTGAA TGCGCAAACC AACCCTTGGC AGAACATATC CATCGCGTCC GCCATCTCCA 3850
GCAGCCGCAC GCGGCGCATC TCGGGCAGCG TTGGGTCCTG GCCACGGGTG CGCATGATCG 3910
TGCTCCTGTC GTTGAGGACC CGGCTAGGCT GGCGGGGTTG CCTTACTGGT TAGCAGAATG 3970
AATCACCGAT ACGCGAGCGA ACGTGAAGCG ACTGCTGCTG CAAAACGTCT GCGACCTGAG 4030
CAACAACATG AATGGTCTTC GGTTTCCGTG TTTCGTAAAG TCTGGAAACG CGGAAGTCAG 4090
CGCCCTGCAC CATTATGTTC CGGATCTGCA TCGCAGGATG CTGCTGGCTA CCCTGTGGAA 4150
CACCTACATC TGTATTAACG AAGCGCTGGC ATTGACCCTG AGTGATTTTT CTCTGGTCCC 4210
GCCGCATCCA TACCGCCAGT TGTTTACCCT CACAACGTTC CAGTAACCGG GCATGTTCAT 4270
CATCAGTAAC CCGTATCGTG AGCATCCTCT CTCGTTTCAT CGGTATCATT ACCCCCATGA 4330
ACAGAAATTC CCCCTTACAC GGAGGCATCA AGTGACCAAA CAGGAAAAAA CCGCCCTTAA 4390
CATGGCCCGC TTTATCAGAA GCCAGACATT AACGCTTCTG GAGAAACTCA ACGAGCTGGA 4450
CGCGGATGAA CAGGCAGACA TCTGTGAATC GCTTCACGAC CACGCTGATG AGCTTTACCG 4510
CAGCTGCCTC GCGCGTTTCG GTGATGACGG TGAAAACCTC TGACACATGC AGCTCCCGGA 4570
GACGGTCACA GCTTGTCTGT AAGCGGATGC CGGGAGCAGA CAAGCCCGTC AGGGCGCGTC 4630
AGCGGGTGTT GGCGGGTGTC GGGGCGCAGC CATGACCCAG TCACGTAGCG ATAGCGGAGT 4690
GTATACTGGC TTAACTATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA CCATATGCGG 4750
TGTGAAATAC CGCACAGATG CGTAAGGAGA AAATACCGCA TCAGGCGCTC TTCCGCTTCC 4810
TCGCTCACTG ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC GAGCGGTATC AGCTCACTCA 4870

FIGURE 5 (cont'd)

AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG CAGGAAAGAA CATGTGAGCA 4930
AAAGGCCAGC AAAAGGCCAG GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTTCCATAGG 4990
CTCCGCCCCC CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG 5050
ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT 5110
CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC CTTCGGGAAG CGTGGCGCTT 5170
TCTCAATGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC 5230
TGTGTGCACG AACCCCCCGT TCAGCCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT 5290
GAGTCCAACC CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT 5350
AGCAGAGCGA GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC TAACTACGGC 5410
TACACTAGAA GGACAGTATT TGGTATCTGC GCTCTGCTGA AGCCAGTTAC CTTCGGAAAA 5470
AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTTGTT 5530
TGCAAGCAGC AGATTACGCG CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT 5590
ACGGGGTCTG ACGCTCAGTG GAACGAAAAC TCACGTTAAG GGATTTTGGT CATGAGATTA 5650
TCAAAAAGGA TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA ATCAATCTAA 5710
AGTATATATG AGTAAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA GGCACCTATC 5770
TCAGCGATCT GTCTATTTCG TTCATCCATA GTTGCCTGAC TCCCCGTCGT GTAGATAACT 5830
ACGATACGGG AGGGCTTACC ATCTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC 5890
TCACCGGCTC CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT 5950
GGTCCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA 6010
AGTAGTTCGC CAGTTAATAG TTTGCGCAAC GTTGTTGCCA TTGCTGCAGG CATCGTGGTG 6070
TCACGCTCGT CGTTTGGTAT GGCTTCATTC AGCTCCGGTT CCCAACGATC AAGGCGAGTT 6130
ACATGATCCC CCATGTTGTG CAAAAAAGCG GTTAGCTCCT TCGGTCCTCC GATCGTTGTC 6190
AGAAGTAAGT TGGCCGCAGT GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT 6250
ACTGTCATGC CATCCGTAAG ATGCTTTTCT GTGACTGGTG AGTAGCTTCA CGCTGCCGCA 6310
AGCACTCAGG GCGCAAGGGC TGCTAAAGGA AGCGGAACAC GTAGAAAGCC AGTCCGCAGA 6370
AACGGTGCTG ACCCCGGATG AATGTCAGCT ACTGGGCTAT CTGGACAAGG GAAAACGCAA 6430
GCGCAAAGAG AAAGCAGGTA GCTTGCAGTG GGCTTACATG GCGATAGCTA GACTGGGCGG 6490

FIGURE 5 (cont'd)

TTTTATGGAC AGCAAGCGAA CCGGAATTGC CAGCTGGGGC GCCCTCTGGT AAGGTTGGGA 6550
AGCCCTGCAA AGTAAACTGG ATGGCTTTCT TGCCGCCAAG GATCTGATGG CGCAGGGGAT 6610
CAAGATCTGA TCAAGAGACA GGATGAGGAT CGTTTCGCAT GATTGAACAA GATGGATTGC 6670
ACGCAGGTTC TCCGGCCGCT TGGGTGGAGA GGCTATTCGG CTATGACTGG GCACAACAGA 6730
CAATCGGCTG CTCTGATGCC GCCGTGTTCC GGCTGTCAGC GCAGGGGCGC CCGGTTCTTT 6790
TTGTCAAGAC CGACCTGTCC GGTGCCCTGA ATGAACTGCA GGACGAGGCA GCGCGGCTAT 6850
CGTGGCTGGC CACGACGGGC GTTCCTTGCG CAGCTGTGCT CGACGTTGTC ACTGAAGCGG 6910
GAAGGGACTG GCTGCTATTG GGCGAAGTGC CGGGGCAGGA TCTCCTGTCA TCTCACCTTG 6970
CTCCTGCCGA GAAAGTATCC ATCATGGCTG ATGCAATGCG GCGGCTGCAT ACGCTTGATC 7030
CGGCTACCTG CCCATTCGAC CACCAAGCGA AACATCGCAT CGAGCGAGCA CGTACTCGGA 7090
TGGAAGCCGG TCTTGTCGAT CAGGATGATC TGGACGAAGA GCATCAGGGG CTCGCGCCAG 7150
CCGAACTGTT CGCCAGGCTC AAGGCGCGCA TGCCCGACGG CGAGGATCTC GTCGTGACTC 7210
ATGGCGATGC CTGCTTGCCG AATATCATGG TGGAAAATGG CCGCTTTTCT GGATTCATCG 7270
ACTGTGGCCG GCTGGGTGTG GCGGACCGCT ATCAGGACAT AGCGTTGGCT ACCCGTGATA 7330
TTGCTGAAGA GCTTGGCGGC GAATGGGCTG ACCGCTTCCT CGTGCTTTAC GGTATCGCCG 7390
CTCCCGATTC GCAGCGCATC GCCTTCTATC GCCTTCTTGA CGAGTTCTTC TGAGCGGGAC 7450
TCTGGGGTTC GAAATGACCG ACCAAGCGAC GCCCAACCTG CCATCACGAG ATTTCGATTC 7510
CACCGCCGCC TTCTATGAAA GGTTGGGCTT CGGAATCGTT TTCCGGGACG CCGGCTGGAT 7570
GATCCTCCAG CGCGGGGATC TCATGCTGGA GTTCTTCGCC CACCCC 7616

Figure 6

```
  NS1
ATGGATCCAAACACTGTGTCAAGCTTTCAGGTAGATTGCTTTCTTTGGCATGTCCGCAAA
------+---------+---------+---------+---------+---------+---
TACCTAGGTTTGTGACACAGTTCGAAAGTCCATCTAACGAAAGAAACCGTACAGGCGTTT

 1
MetAspProAsnThrValSerSerPheGlnValAspCysPheLeuTrpHisValArgLys

CGAGTTGCAGACCAAGAACTAGGTGATGCCCCATTCCTTGATCGGCTTCGCCGAGATCAG
------+---------+---------+---------+---------+---------+---
GCTCAACGTCTGGTTCTTGATCCACTACGGGGTAAGGAACTAGCCGAAGCGGCTCTAGTC

ArgValAlaAspGlnGluLeuGlyAspAlaProPheLeuAspArgLeuArgArgAspGln

AAATCCCTAAGAGGAAGGGGCAGCACTCTTGGTCTGGACATCGAGACAGCCACACGTGCT
------+---------+---------+---------+---------+---------+---
TTTAGGGATTCTCCTTCCCCGTCGTGAGAACCAGACCTGTAGCTCTGTCGGTGTGCACGA

LysSerLeuArgGlyArgGlySerThrLeuGlyLeuAspIleGluThrAlaThrArgAla

GGAAAGCAGATAGTGGAGCGGATTCTGAAAGAAGAATCCGATGAGGCACTTAAAATGACC
------+---------+---------+---------+---------+---------+---
CCTTTCGTCTATCACCTCGCCTAAGACTTTCTTCTTAGGCTACTCCGTGAATTTTACTGG

GlyLysGlnIleValGluArgIleLeuLysGluGluSerAspGluAlaLeuLysMetThr

                 HA2
ATGCAGATCCCGGCTGTGGGTAAAGAATTCAACAAATTAGAAAAAAGGATGGAAAATTTA
------+---------+---------+---------+---------+---------+---
TACGTCTAGGGCCGACACCCATTTCTTAAGTTGTTTAATCTTTTTTCCTACCTTTTAAAT

81  linker   65          69
MetGlnIleProAlaValGlyLysGluPheAsnLysLeuGluLysArgMetGluAsnLeu

AATAAAAAAGTTGATGATGGATTTCTGGACATTTGGACATATAATGCAGAATTGTTAGTT
------+---------+---------+---------+---------+---------+---
TTATTTTTTCAACTACTACCTAAAGACCTGTAAACCTGTATATTACGTCTTAACAATCAA

 81
AsnLysLysValAspAspGlyPheLeuAspIleTrpThrTyrAsnAlaGluLeuLeuVal

CTACTGGAAAATGAAAGGACTCTGGATTTCCATGACTCAAATGTGAAGAATCTGTATGAG
------+---------+---------+---------+---------+---------+---
GATGACCTTTTACTTTCCTGAGACCTAAAGGTACTGAGTTTACACTTCTTAGACATACTC

LeuLeuGluAsnGluArgThrLeuAspPheHisAspSerAsnValLysAsnLeuTyrGlu
```

Figure 6 (Cont'd.)

```
AAAGTAAAAAGCCAATTAAAGAATAATGCCAAAGAAATCGGAAATGGATGTTTTGAGTTC
------+---------+---------+---------+---------+---------+---
TTTCATTTTTCGGTTAATTTCTTATTACGGTTTCTTTAGCCTTTACCTACAAAACTCAAG

LysValLysSerGlnLeuLysAsnAsnAlaLysGluIleGlyAsnGlyCysPheGluPhe

TACCACAAGTGTGACAATGAATGCATGGAAAGTGTAAGAAATGGGACTTATGATTATCCC
------+---------+---------+---------+---------+---------+---
ATGGTGTTCACACTGTTACTTACGTACCTTTCACATTCTTTACCCTGAATACTAATAGGG

                                150
TyrHisLysCysAspAsnGluCysMetGluSerValArgAsnGlyThrTyrAspTyrPro

AAATATTCAGAAGAGTCAAAGTTGAACAGGGAAAAGGTAGATGGAGTGAAATTGGAATCA
------+---------+---------+---------+---------+---------+---
TTTATAAGTCTTCTCAGTTTCAACTTGTCCCTTTTCCATCTACCTCACTTTAACCTTAGT

LysTyrSerGluGluSerLysLeuAsnArgGluLysValAspGlyValLysLeuGluSer

ATGGGGATCTATCAGATTCTGGCGATCTACTCAACTGTCGCCAGTTCACTGGTGCTTTTG
------+---------+---------+---------+---------+---------+---
TACCCCTAGATAGTCTAAGACCGCTAGATGAGTTGACAGCGGTCAAGTGACCACGAAAAC

MetGlyIleTyrGlnIleLeuAlaIleTyrSerThrValAlaSerSerLeuValLeuLeu

GTCTCCCTGGGGGCAATCAGTTTCTGGATGTGTTCTAATGGATCTTTGCAGTGCAGAATA
------+---------+---------+---------+---------+---------+---
CAGAGGGACCCCCGTTAGTCAAAGACCTACACAAGATTACCTAGAAACGTCACGTCTTAT

ValSerLeuGlyAlaIleSerPheTrpMetCysSerAsnGlySerLeuGlnCysArgIle

TGCATCTGA
------+--
ACGTAGACT

   222
CysIle
```

FIGURE 7

GGC ATA TTC GGC GCA ATA GCA GGT TTC ATA GAA AAT GGT 39
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly
1 5 10

TGG GAG GGA ATG ATA GAC GGT TGG TAC GGT TTC AGG CAT 78
Trp Glu Gly Met Ile Asp Gly Trp Tyr Gly Phe Arg His
15 20 25

CAA AAT TCT GAG GGC ACA GGA CAA GCA GCA GAT CTT AAA 117
Gln Asn Ser Glu Gly Thr Gly Gln Ala Ala Asp Leu Lys
30 35

AGC ACT CAA GCA GCC ATC GAC CAA ATC AAT GGG AAA CTG 156
Ser Thr Gln Ala Ala Ile Asp Gln Ile Asn Gly Lys Leu
40 45 50

AAT AGG GTA ATC GAG AAG ACG AAC GAG AAA TTC CAT CAA 195
Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His Gln
55 60 65

ATC GAA AAG GAA TTC TCA GAA GTA GAA GGG AGA ATT CAG 234
Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln
70 75

GAC CTC GAG AAA TAC GTT GAA GAC ACT AAA ATA GAT CTC 273
Asp Leu Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu
80 85 90

TGG TCT TAC AAT GCG GAG CTT CTT GTC GCT CTG GAG AAC 312
Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu Asn
95 100

CAA CAT ACA ATT GAT CTG ACT GAC TCG GAA ATG AAC AAA 351
Gln His Thr Ile Asp Leu Thr Asp Ser Glu Met Asn Lys
105 110 115

CTG TTT GAA AAA ACA AGG AGG CAA CTG AGG GAA AAT GCT 390
Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu Asn Ala
120 125 130

GAG GAC ATG GGC AAT GGT TGC TTC AAA ATA TAC CAC AAA 429
Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys
135 140

TGT GAC AAT GCT TGC ATA GGG TCA ATC AGA AAT GGG ACT 468
Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr
145 150 155

TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA AAC AAC 507
Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn
160 165

CGG TTT CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA TAC 546
Arg Phe Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr
170 175 180

FIGURE 7 (cont'd)

```
AAA GAC TGG ATC CTG TGG ATT TCC TTT GCC ATA TCA TGC        585
Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala Ile Ser Cys
        185                     190                 195

TTT TTG CTT TGT GTT GTT TTG CTG GGG TTC ATC ATG TGG        624
Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
                200                     205

GCC TGC CAG AAA GGC AAC ATT AGG TGC AAC ATT TGC ATT        663
Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                     215                     220

TGA                                                        666
```

FIGURE 8

```
GGC ATA TTC GGC GCA ATA GCA GGT TTC ATA GAA AAT GGT        39
Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly
 1               5                  10

TGG GAG GGA ATG ATA GAC GGT TGG TAC GGT TTC AGG CAT        78
Trp Glu Gly Met Ile Asp Gly Trp Tyr Gly Phe Arg His
    15                  20                  25

CAA AAT TCC GAG GGC ACA GGA CAA GCA GCA GAT CTT AAA       117
Gln Asn Ser Glu Gly Thr Gly Gln Ala Ala Asp Leu Lys
                30                  35

AGC ACT CAA GCA GCC ATC GAC CAA ATC AAT GGG AAA CTG       156
Ser Thr Gln Ala Ala Ile Asp Gln Ile Asn Gly Lys Leu
 40                  45                  50

AAT AGG GTA ATC GAG AAG ACG AAC GAG AAA TTC CAT CAA       195
Asn Arg Val Ile Glu Lys Thr Asn Glu Lys Phe His Gln
        55                  60                  65

ATC GAA AAG GAA TTC TCA GAA GTA GAA GGG AGA ATT CAG       234
Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln
                    70                  75

GAC CTC GAG AAA TAC GTT GAA GAC ACT AAA ATA GAT CTC       273
Asp Leu Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu
    80                  85                  90

TGG TCT TAC AAT GCG GAG CTT CTT GTC GCT CTG GAG AAC       312
Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Leu Glu Asn
                95                 100

CAA CAT ACA ATT GAT CTG ACT GAC TCG GAA ATG AAC AAA       351
Gln His Thr Ile Asp Leu Thr Asp Ser Glu Met Asn Lys
105                 110                 115

CTG TTT GAA AAA ACA AGG AGG CAA CTG AGG GAA AAT GCT       390
Leu Phe Glu Lys Thr Arg Arg Gln Leu Arg Glu Asn Ala
        120                 125                 130

GAG GAC ATG GGC AAT GGT TGC TTC AAA ATA TAC CAC AAA       429
Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys
                    135                 140

TGT GAC AAT GCT TGC ATA GGG TCA ATC AGA AAT GGG ACT       468
Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr
    145                 150                 155

TAT GAC CAT GAT GTA TAC AGA GAC GAA GCA TTA AAC AAC       507
Tyr Asp His Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn
                160                 165
```

FIGURE 8 (cont'd)

```
CGG TTT CAG ATC AAA GGT GTT GAA CTG AAG TCA GGA TAC        546
Arg Phe Gln Ile Lys Gly Val Glu Leu Lys Ser Gly Tyr
170                 175                 180

AAA GAC TGG ATC CTG TGG ATT TCC TTT GCC ATA TCA TGC        585
Lys Asp Trp Ile Leu Trp Ile Ser Phe Ala Ile Ser Cys
        185                 190                 195

TTT TTG CTT TGT GTT GTT TTG CTG GGG TTC ATC ATG TGG        624
Phe Leu Leu Cys Val Val Leu Leu Gly Phe Ile Met Trp
                200                 205

GCC TGC CAA AAA GGC AAC ATT AGG TGC AAC ATT TGC ATT        663
Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
    210                 215                 220

TGA                                                        666
```

INTERNATIONAL SEARCH REPORT

International application No. PCT/US94/01149

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :A61K39/12; C12P 21/06, C12N 15/00; C07K 3/00, 13/00, 15/00, 17/00
US CL : 424/89; 530/350; 435/69.1, 172.1, 320.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/89; 530/350; 435/69.1, 172.1, 320.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Dialog, search terms: influenza, NS1, HA, vaccine, recombinant DNA, polypeptide

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EPA 0,366,238 (YOUNG, et al), 02 May 1990, pages 1-25, see pages 3 and 7.	1-6, 11-16, 21-23, 27-30
X	The Journal of Immunology, Volume 140, Number 4, issued 15 February 1988, Kuwano et al, "HA2 Subunit of Influenza A H1 and H2 Subtype Viruses Induces A Protective Cross-Reactive Cytotoxic T Lymphocyte Response", pages 1264-1268, see entire document.	1-6, 11-16
X	Virology, Volume 178, issued 1990, Kuwano et al, "Cross-Reactive Protection Against Influenza A Virus Infections by an NS1-Specific CTL Clone", pages 174-179, see entire document.	1-3, 11-13, 21, 22, 28, 29

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 MAY 1994	JUN 01 1994

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	LYNETTE F. SMITH Jill Warden for
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

Int ational application No.
PCT/US94/01149

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-6, 11-16,21-23, 27-30

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

I. Claims 1-6, 11-16, 21-23 and 27-30, drawn to a vaccine containing an immunogenic fragment of Type A influenza virus and including one polypeptide, protein, microorganism and vector, classes 424,530 and 435, subclasses, 89, 350 and 69.1, 172.1, 320.1.

II. Claims 1-6, 11-16, 21-23 and 27-30, drawn to a vaccine containing an immunogenic fragment of Type B influenza virus and including one polypeptide, protein, microorganism and vector, classes 424,530 and 435, subclasses, 89, 350 and 69.1, 172.1, 320.1.

III. Claims 1-5, 7, 11-15, 17, 21, 22, 25 and 27-29, drawn to a vaccine containing an immunogenic fragment of Type A influenza virus and including one polypeptide, protein, microorganism and vector, classes 424,530 and 435, subclasses, 89, 350 and 69.1, 172.1, 320.1.

IV. Claims 1-5, 7, 11-15, 17, 21, 22, 25 and 27-29, drawn to a vaccine containing an immunogenic fragment of Type B influenza virus and including one polypeptide, protein, microorganism and vector, classes 424,530 and 435, subclasses, 89, 350 and 69.1, 172.1, 320.1.

V. Claims 1-5, 8, 11-15, 18, 21, 22, 25 and 27-29, drawn to a vaccine containing an immunogenic fragment of Type A influenza virus and including onepolypeptide, protein, microorganism and vector, classes 424, 530 and 435, subclasses, 89, 350 and 69.1, 172.1, 320.1.

VI. Claims 1-5, 8, 11-15, 18, 21, 22, 24 and 27-29 drawn to a vaccine containing an immunogenic fragment of Type B influenza virus and including one polypeptide, protein, microorganism and vector, classes 424,530 and 435, subclasses, 89, 350 and 69.1, 172.1, 320.1.

VII. Claims 1-5, 9, 11-15, 19, 21, 22, 26-29 and 31, drawn to a vaccine containing an immunogenic fragment of Type A influenza virus and including one polypeptide, protein, microorganism and vector, classes 424,530 and 435, subclasses, 89, 350 and 69.1, 172.1, 320.1.

VIII. Claims 1-5, 9, 11-15, 19, 21, 22, 26-29 and 31 drawn to a vaccine containing an immunogenic fragment of Type B influenza virus and including one polypeptide, protein, microorganism and vector, classes 424,530 and 435, subclasses, 89, 350 and 69.1, 172.1, 320.1.

IX. Claims 1-5, 11-15, 20-22, 27-29 and 32, drawn to a vaccine containing an immunogenic fragment of Type A influenza virus and including one polypeptide, protein, microorganism and vector, classes 424,530 and 435, subclasses, 89, 350 and 69.1, 172.1, 320.1.

X. Claims 1-5, 11-15, 20-22, 27-29 and 32, drawn to a vaccine containing an immunogenic fragment of TypeBinfluenzavirusand including one polypeptide, protein, microorganism and vector, classes 424,530 and 435, subclasses, 89, 350 and 69.1, 172.1, 320.1.

XI. Claims 10, 33, 34 and 35, drawn to a vaccine containing two polypeptides, wherein the first polypeptide has SEQ ID No. 10.

XII. Claims 33, 34 and 35, drawn to a vaccine containing two polypeptides, wherein the first polypeptide has SEQ ID No. 12.

XIII. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 16.

XIV. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 18.

XV. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 20.

XVI. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 22.

XVII. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 24.

XVIII. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No.26.

XIX. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 27.

XX. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 28.

XXI. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 29.

XXII. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 30.

XXIII. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 32.

XXIV. Claims 33, 35 and 36, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 34.

XXV. Claims 33, 35 and 37, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 14.

XXVI. Claims 33, 35 and 38, drawn to a vaccine containing two polypeptides, wherein the second polypeptide has SEQ ID No. 57.

XXVII. Claims 39 and 40 drawn to a vaccine containing three polypeptides.

and it considered that the International Application did not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The claims of groups I-XXVII are drawn to multiple products which are not linked by a special technical feature so as to form a single inventive concept. PCT Rule 13.1 and Rule 13.2 do not provide for multiple products and methods within a single general inventive concept.